A Quantitative Landauer's Principle

Philippe Faist,[1] Frédéric Dupuis,[1] Jonathan Oppenheim,[2] and Renato Renner[1]

[1] *Institute for Theoretical Physics, ETH Zurich, 8093 Switzerland*

[2] *Department for Physics and Astronomy, University College of London, WC1E 6BT London, U.K.*

(Dated: November 7, 2012)

Landauer's Principle states that the work cost of erasure of one bit of information has a fundamental lower bound of $kT\ln(2)$. Here we prove a quantitative Landauer's principle for arbitrary processes, providing a general lower bound on their work cost. This bound is given by the minimum amount of (information theoretical) entropy that has to be dumped into the environment, as measured by the conditional max-entropy. The bound is tight up to a logarithmic term in the failure probability. Our result shows that the minimum amount of work required to carry out a given process depends on how much correlation we wish to retain between the input and the output systems, and that this dependence disappears only if we average the cost over many independent copies of the input state. Our proof is valid in a general framework that specifies the set of possible physical operations compatible with the second law of thermodynamics. We employ the technical toolbox of matrix majorization, which we extend and generalize to a new kind of majorization, called *lambda-majorization*. This allows us to formulate the problem as a semidefinite program and provide an optimal solution.

arXiv:1211.1037v1 [quant-ph] 5 Nov 2012

Introduction.—Landauer's Principle [1, 2], and more generally the relation between the second law of thermodynamics and information theory, has received much attention in the past decades. Studies have notably focused on fundamental limits on heat generated by computation [2], the exorcism of Maxwell's demon via information theory (see eg. [3]), and generalizations to quantum settings such as the characterization of entanglement through thermodynamical considerations [4], or the determination of the work cost of information erasure with the help of quantum side information [5].

Landauer's Principle can be stated in the following way. Consider the erasure process of an unknown bit, i.e. the logical operation that resets the bit to a reference state (e.g. zero). Landauer's Principle asserts that any physical implementation that performs this erasure, using a heat bath at temperature T, has a work cost of at least $kT\ln(2)$, where k is the Boltzmann constant. More generally, Landauer noted that all irreversible operations, and not only the erasure of a bit, must cost work due to the transfer of entropy from the information-bearing degrees of freedom to the environment, which causes the system to dissipate heat. Bennett refined the formulation of this principle and showed its relevance in thermodynamics (exorcising the Maxwell demon [2, 3]) and in computation [6].

The work cost of thermodynamic processes in the context of information theory has been studied for various classical and quantum systems. Szilard [7] originally considered a single-particle gas enclosed in a box with a piston and noted that $kT\ln(2)$ work could be reversibly extracted from the gas at the expense of losing the information about which side of the piston the particle is on. The reverse process corresponds to erasing this information, bringing the particle on one definite side at $kT\ln(2)$ work cost. Landauer [1, 8] studied the example of a particle in a double-V shaped potential, which represents a bit of information, and showed that its erasure costs work. While these results apply to fully unknown bits, the bounds have to be adapted if the system we erase is partially known. In such a case, the average amount of work needed is lower bounded by $kT\ln(2)\,H(X)$, where $H(X)$ is the Shannon entropy of the system X and where the average is taken over many independent repetitions of the erasure process [3, 9, 10]. This result has been derived and extended in several contexts such as using quantum computers performing data compression [11], Hamiltonian models [12] or in a resource theory framework [13, 14]. This bound can also be generalized to other processes, for which the average work cost is then given by the amout of entropy the processes transfers into the environment. We refer to Janzing [15, 16] for a proof in a resource theory framework.

Generalizations to a single-shot regime, where statements are made about individual processes rather than many repetitions of them, have been proposed, for example in terms of majorization conditions [13], and in terms of entropic quantities which take into account approximate transitions and a probability of failure [17–19]. Explicit Hamiltonian models have also been used to study the case of erasure with quantum side information [5]. It is usually assumed that the system carrying the information has a degenerate Hamiltonian. More recently, these thermodynamic considerations have been extended to the case of non-degenerate Hamiltonians [19–21], and the majorization condition also adapted to this scenario [17, 19, 21], based on ideas from [22–25].

In the present article, we revisit Landauer's principle in the light of general quantum processes. Our main result is an explicit and rigourous expression for the fundamental minimal work cost of any process $\mathcal{E}$ that acts on a system X and brings it from a state σ to a new state ρ. The bound is robust, i.e. it holds even if one tolerates an error probability ε. The work cost W of such a process is lower bounded by the amount of entropy that has to be dumped into the environment, as measured by the

smooth conditional max-entropy [26, 27],

$$W \geqslant kT \ln(2)\, H_{\max}^{\varepsilon} \left(E|X\right)_{\rho} \ . \tag{1}$$

Here, the entropy is evaluated for the state ρ which is a purification of the output state obtained by applying the process $\mathcal{E}$ to a purification of the input state σ_X (see Proposition 3). The entropy measure, $H_{\max}^{\varepsilon}$, is part of the smooth entropy framework that is widely used in single-shot quantum information theory [26–30]. Its formal definition will be given later.

Our quantitative Landauer's Principle is tight up to logarithmic terms in the failure probability of the implementation of the process $\mathcal{E}$. Indeed, we can devise an explicit process carrying out the requested mapping $\mathcal{E}$ that is nearly optimal. This near-optimal process is based on the scheme proposed by del Rio *et al.* [5], which erases a system using available quantum side information.

Our bound is valid in a general framework that specifies the set of physically allowed operations. This framework conceptually separates the operations that are intrinsically thermodynamical (e.g., the erasure of information) from those that simply correspond to reversible information processing (e.g., unitaries and the addition of ancillas). The former will be those that cost work or that are capable of extracting work from a system; the latter are done for free, i.e. at no work cost. We assume that our systems have a completely degenerate Hamiltonian. The set of allowed operations is motivated by the second law of thermodynamics, which forbids cyclic processes whose net effect is to extract work.

For the proofs we use a characterization of our framework by a relatively simple and intuitive generalization of the notion of majorization which is inspired by previous work where the eigenvalues of the input are rescaled until the input majorizes the output [17, 19], achieved for example by appending a work system [21]. We term our generalisation *lambda-majorization*, and provide a mathematical characterization of this notion in terms of completely positive maps that satisfy some normalization conditions.

In the asymptotic limit of many identical and identically distributed (i.i.d.) copies of these systems (i.e., the process is repeated $n \to \infty$ independent times, $\mathcal{E}^{\otimes n}$, on n i.i.d. input states $\sigma^{\otimes n}$), we obtain as a corollary of our main result a value for the average work cost of erasure per copy,

$$\langle W \rangle \geqslant \left[H\left(X\right)_{\sigma} - H\left(X\right)_{\rho}\right] kT \ln(2) \ , \tag{2}$$

which is in agreement with the informal formulation of Landauer's principle, that the work cost of any process is determined by the decrease of entropy in the information-bearing degrees of freedom (see [16] for a proof in a resource framework).

We should point out that the general bound (1) can be arbitrarily larger than the average bound (2). This deviation highlights an important feature, namely that correlations between the input and the output of the transformation play a significant role in the single-shot regime. It is important to not only consider the input and output states, but also the whole process, or computation, that is performed on the actual input. This is natural and generalizes the classical case where this consideration is obvious, since a classical computer acts on the actual state of a register and not on its probability distribution. In the quantum case, we specify the full algorithm (or computation) as a completely positive map, which inherently tells us which correlations are preserved between the input and output systems. While the transformation of a state into another (e.g. in a resource theoretic approach) is a relevant question, we focus in this paper on the case where the computation is given, thus fixing all the correlations that are preserved or destroyed between the input and the output.

As a simple example, consider X to be a fully mixed qubit, i.e. in the state $\sigma_X = \frac{1}{2}\mathbb{1}_2$. Suppose we wish to transform this state into another fully mixed qubit again, $\rho_X = \frac{1}{2}\mathbb{1}_2$. There are two obvious processes that achieve this goal: we may (a) simply copy the input qubit to the output, or (b) throw away the input and prepare a new fully mixed qubit. Both processes (a) and (b) provide the required output. However, if we had information about the specific state in which the qubit initially was (e.g. suppose we had kept a qubit C that was maximally entangled with the input), then in the case of process (a), C would remain entangled with the output; however in the case of (b), C would have lost all correlations with the output qubit. In this first example, both processes cost no work: (a) is the identity process, and in (b), the work dissipated to erase the qubit is retrieved again when we prepare a new mixed qubit.

However, the work costs of these processes differ if we consider less trivial input and output states. Let X be a quantum system composed of $n+1$ qubits, in a state σ_X where the first qubit is randomly zero or one with probability $1/2$, and the n remaining qubits are either all zero if the first qubit is zero, or all in a fully mixed state if the first qubit is one. This state has the distribution $\{1/2, 2^{-(n+1)}, 2^{-(n+1)}, \ldots 2^{-(n+1)}\}$ and is depicted in Figure 1. Suppose that we wish to bring this system into the state $\rho_X = \sigma_X$, i.e. the same state as the input state, using either process (a) or (b) again. Process (a) would simply copy the input to its output, and would not cost any work, since it is the identity channel. However, process (b) first has to erase the input state and then prepare the output state. If we are lucky, the n qubits are in state $|0\ldots0\rangle\langle0\ldots0|$ (if the first qubit is $|0\rangle$) and we can just erase the first qubit using $kT\ln(2)$ work. However, if we want to erase the system with certainty, we have to consider the worst case in which we have to erase n fully mixed qubits (which occurs with the non-negligible probability $1/2$). So the erasure work cost may be as bad as $(n+1)kT\ln(2)$. In order to prepare this state again as the output of the process, we may think of tossing a coin to decide in which state $|0\ldots0\rangle\langle0\ldots0|$ or $|1\rangle\langle1| \otimes 2^{-n}\mathbb{1}_{2^n}$ to prepare X in. If we are lucky, we have to prepare a

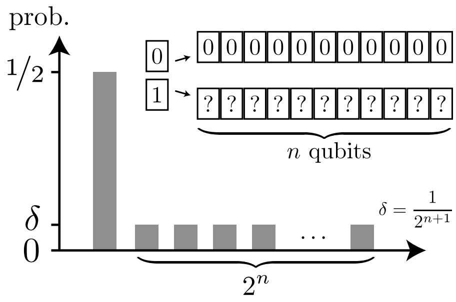

FIG. 1: The probability distribution of a state in which single-shot effects become important, even for large systems. A register of $n+1$ qubits are in a state ρ such that if the first qubit is zero (with probability $1/2$), then all the rest are zero too; if the first qubit is one (with probability $1/2$), then all the rest are in a fully mixed state. The spectrum has a large eigenvalue ($1/2$), but also has a large support size (2^n+1); as a consequence, $H^{\varepsilon}_{\min}(\rho) \approx 1$ and $H^{\varepsilon}_{\max}(\rho) \approx n$ can differ by an arbitrary amount.

mixed state on n qubits and extract $nkT\ln(2)$ work in the process, but in the worst case, we have to prepare $|0\dots0\rangle\langle0\dots0|$ and can't extract more than just $kT\ln(2)$ (from the coin toss). Hence, in the worst case, process (b) costs a total of $n\,kT\ln(2)$ work, which can be arbitrarily larger than the (zero) cost of process (a); in fact, the gap diverges as $n \to \infty$.

This example shows that in the general single-shot regime, the specification of only the input state σ_X and the output state ρ_X does not suffice, and correlations between the input and the output contribute to determine the minimal work cost of the process (although these correlations are not relevant in the asymptotic i.i.d. regime). Our result (1) incorporates this property intrinsically and provides a bound that is valid for any given process.

The remainder of this paper is organized as follows. We will first present the mathematical framework used to model thermodynamic processes. We then introduce lambda-majorization, which captures all possible operations in our framework. Lambda-majorization is characterized in terms of completely positive maps that satisfy some specific normalization conditions, and we use this characterization to derive the main result by formulating the problem as a semidefinite program. The latter is solved by providing optimal primal and dual feasible plans with the same value, which guarantees optimality of the result. Finally, some special cases are derived which recover some previously known results.

Framework.—Consider a quantum mechanical system X in an inital state described by the density operator σ. Our task is to bring the system X to another state ρ, while attempting to maximize some kind of notion of "extracted" work in the process. Throughout this paper we assume that the Hamiltonians of the systems we consider are completely degenerate.

We first postulate two basic operations of thermodynamical nature, involving a heat bath at temperature T: the erasure of a single qubit to a pure state at $kT\ln(2)$ work cost, and the corresponding reverse process which extracts $kT\ln(2)$ work by transforming a pure state into a fully mixed state. Here k is the Boltzmann constant. These operations are motivated by the variety of explicit physical thermodynamical frameworks in which they can be performed, for example using Szilard boxes [7, 18] or by isothermally manipulating energy levels of Hamiltonians [5, 12, 20]. Crucially, we assume the second law of thermodynamics, and require that there exist no operation that would allow us to form a cycle for which the net effect would be the extraction of work. This justifies that no other work extraction procedure can yield more work than $kT\ln(2)$ from a pure qubit, or else a cycle with net work gain could be formed by appending an erasure process, itself only costing $kT\ln(2)$.

Apart from this constraint on the set of allowed operations, it is natural to also allow usual quantum information processing. Since our Hamiltonians are degenerate, we can allow all global unitaries and they cost no work. We do not need to use the fact that these unitaries are implementable by a device operating in contact with a heat bath, since expanding the class of allowable operations actually strengthens the bound we derive. In practice, one has very crude local control over the operations, and the acting agent does not know which unitary is being implemented, however, this is actually not an obstacle for implementation [11, 14]. In addition to unitaries, we will allow pure ancillas to be added to the system, which permits more general computation. Crucially, ancillas will have to be restored to their initial pure state, so that it is not possible to "hide" a work cost in an ancilla that was left mixed.

The following framework is motivated by the above considerations. The processes we allow are (finite) combinations of the following elementary operations:

(a) Bring n qubits (of the system X or an ancilla A) from any state to a pure state ('erasure') at cost $n\,kT\ln 2$ work;

(b) Bring n qubits (of the system X or an ancilla A) from a pure state to a fully mixed state while extracting $n\,kT\ln 2$ work;

(c) Add and remove ancillas in a pure state at no work cost, as long as all the ancillas have been restored to their initial pure state at the end of the process;

(d) Perform arbitrary unitaries (over X and any added ancillas) at no work cost.

Operations (a) and (b) are those of thermodynamical nature, and may be carried out in a wide range of existing frameworks as mentioned above. One may view these operations as *defining* a quantity which we call "work".

On the other hand, operations (c) and (d) are purely information-theoretical. They allow us to perform any quantum information processing circuit, since we allow pure ancillas to be added. However, there is the condition that "randomness" may not be disposed of for free,

namely that ancillas have to be restored to their initial pure states at the end of the process.

Lambda-Majorization.—We will now provide a simple mathematical characterization of all operations allowed in our framework.

First, note that the operations (a)–(d) allow the use of so-called *noisy operations* [13], which correspond to adding an ancilla system N in a fully mixed state, performing a joint unitary, and removing the ancilla. Specifically, a noisy operation is composed in our framework of first an operation of type (c) (adding a pure ancilla of n qubits), followed by an operation of type (b) (extracting $n\,kT\ln 2$ work from the ancilla making it fully mixed), then one of type (d) (performing the necessary unitary to carry out the noisy operation), and finally an operation of type (a) (erasing the ancilla back to its pure state at a work cost $n\,kT\ln 2$). The total process has a work balance of zero. This means that we may thus carry out noisy operations for free within our framework and use them as building blocks for more complex processes. In the following, we deal implicitly with the ancilla N and it should not be confused with further ancillas that will be added.

The following result by Horodecki *et al.* [13] relates noisy operations to the mathematical notion of majorization [31, 32].

Noisy Operations and Majorization. *The transition on system X from state σ to state ρ is possible by noisy operation if and only if $\sigma \succ \rho$.*

Majorization between two (normalized) states $\sigma \succ \rho$ captures the fact that ρ is "more mixed" than σ, or that the eigenvalues of ρ can be written as a "mixture" of the eigenvalues of σ. Formally, majorization can be characterized by the existence of a unital, trace-preserving completely positive map that brings σ to ρ [33–36]. A channel $\mathcal{E}$ is *trace-preserving* if $\mathcal{E}^\dagger(\mathbb{1}) = \mathbb{1}$ and *unital* if $\mathcal{E}(\mathbb{1}) = \mathbb{1}$.

Proposition 1. *Two positive matrices σ and ρ satisfy $\sigma \succ \rho$ if and only if there exists a trace-preserving, unital, completely positive map $\mathcal{E}$ satisfying $\mathcal{E}(\sigma) = \rho$.*

The notion of majorization is discussed in more detail in Appendix A.

We will now provide some background insight for the meaning of our new concept of lambda-majorization. The idea is to characterize "how well" a state σ majorizes a state ρ. Suppose that we have a system X in state σ_X and we want to bring it to the state ρ_X, where $\sigma_X \succ \rho_X$. In this case, one can simply carry out a noisy operation as described above. Suppose now that we have an ancilla A that is in a fully mixed state, $\frac{\mathbb{1}_A}{|A|}$, and suppose that we are fortunate enough for $\sigma_X \otimes \frac{\mathbb{1}_A}{|A|} \succ \rho_X \otimes |0\rangle\langle 0|_A$ to also hold (for some pure state $|0\rangle_A$ on A). Then by applying a joint noisy operation on both systems, this would correspond to actually erasing the system A "for free" during the transition $\sigma \to \rho$. We could then say that the randomness of the ancilla A was "transferred" into system X. We will view this type of transition as *work extraction* on system X during a transition $\sigma_X \to \rho_X$.

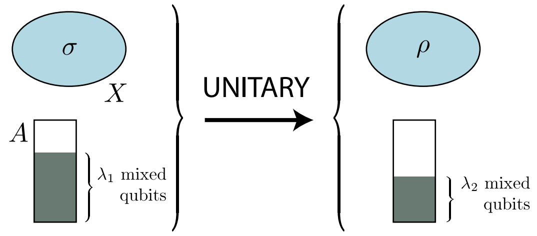

FIG. 2: Lambda-Majorization corresponds to absorbing a certain amount of randomness from an ancilla during a unitary operation. The system X starts in state σ, and the ancilla A in a state with λ_1 fully mixed qubits with the remaining qubits pure. The goal is to devise a global unitary that will bring the system X to the state ρ, while leaving the least possible number λ_2 of fully mixed qubits in A. The difference $\lambda = \lambda_1 - \lambda_2$, is the work extracted by the process; if the value is negative, it corresponds to a work cost. In the main text, we allow a noisy operation instead of a unitary operation, but one could simply add more mixed qubits to the ancilla on each side and use those to implement a noisy operation with a unitary.

In another situation, it might be that $\sigma_X \not\succ \rho_X$. However, in that case, for a large enough ancilla A the majorization $\sigma_X \otimes |0\rangle\langle 0|_A \succ \rho_X \otimes \frac{\mathbb{1}_A}{|A|}$ will hold. The corresponding noisy operation then leaves us with a mixed ancilla that started off pure; we will view such a transition on system X as *costing work.*

Such operations can be performed within our framework, using operations (a)–(d). In particular, the relation to work is given by elementary erasure and work extraction (operations (a) and (b)) applied to the ancilla A after the transition to restore it to its initial state.

In general, the ancilla A may start with λ_1 mixed qubits and end up with λ_2 mixed qubits after a noisy operation; we consider in this case to have extracted $(\lambda_1 - \lambda_2)\,kT\ln(2)$ amount of work. This situation is depicted in Figure 2. Both considerations above about work cost and work extraction are encompassed, simply because we count the difference in the "amount of randomness" present in the ancilla before and after the process. This is the idea behind the concept of lambda-majorization, whose definition we can now state.

Lambda-Majorization. *For two density operators σ_X, ρ_Y on two systems X and Y, we will say that σ_X λ-majorizes ρ_Y, denoted by $\sigma_X \xrightarrow{\lambda} \rho_Y$, if there exists a (large enough) ancilla system A, as well as $\lambda_1, \lambda_2 \geqslant 0$ with $\lambda = \lambda_1 - \lambda_2$, such that*

$$2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}} \otimes \sigma_X \succ 2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}} \otimes \rho_X \ ,$$

where $2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}}$ and $2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}}$ are fully mixed states on λ_1 (respectively λ_2) qubits of A, and where the remaining qubits of A in each case are pure.

An expression for "by how much" a state majorizes another was originally introduced in [17] and used in [19], in the context of work extraction games from Szilard boxes. Their measure, the "relative mixedness" between σ and ρ, corresponds to the optimal λ such that $\sigma \xrightarrow{\lambda} \rho$.

Lambda-majorization captures the possible processes that are allowed in our framework. Indeed, if $\sigma \xrightarrow{\lambda} \rho$, then one has $2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}} \otimes \sigma \succ 2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}} \otimes \rho$ for some λ_1, λ_2 with $\lambda = \lambda_1 - \lambda_2$. Hence, there exists a noisy operation (itself a combination of operations (a)–(d) with zero total work cost) that performs the transition from $2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}} \otimes \sigma$ to $2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}} \otimes \rho$. The λ_1 mixed qubits that we have appended to σ can be created by appending a large pure ancilla (operation (c)), and using operation (b) to extract $\lambda_1\, kT\ln(2)$ work from λ_1 qubits, rendering them fully mixed. At the end of the process, after the noisy operation, we need to restore the ancilla in a pure state; we thus need to erase (operation (a)) the remaining λ_2 qubits, costing $\lambda_2\, kT\ln(2)$ work. The total extracted work is then $(\lambda_1 - \lambda_2)\, kT\ln(2) = \lambda\, kT\ln(2)$. Conversely, each individual operation (a)–(d), individually transforming some state σ into a state ρ and costing work W, implies the lambda-majorization $\sigma \xrightarrow{\lambda} \rho$ with $W = -\lambda\, kT\ln(2)$. This is clear for operations (c) and (d). For operations (a) and (b), this follows from results derived in Appendix A 3.

The ancilla system above may be viewed as some kind of "information battery", as was suggested by Bennett [2] who suggested using a blank memory tape as "fuel" to extract work. In this case, the ancilla can be used as a storage of "purity" (or as a storage for "mixedness" or "randomness" which we would like to get rid of), which is increased or decreased by processes like the ones suggested above.

It turns out that one can characterize lambda-majorization by the existence of a completely positive map satisfying some special normalization conditions, analogously to Proposition 1.

Proposition 2. *Two normalized density matrices σ_X and ρ_Y on two systems X and Y satisfy $\sigma_X \xrightarrow{\lambda} \rho_Y$ if and only if there exists a completely positive map $\mathcal{T}_{X\to Y}$ satisfying $\rho_Y = \mathcal{T}_{X\to Y}(\sigma_X)$, such that $\mathcal{T}^\dagger_{X\to Y}(\mathbb{1}_X) \leqslant \mathbb{1}_Y$ and $\mathcal{T}_{X\to Y}(\mathbb{1}_X) \leqslant 2^{-\lambda}\mathbb{1}_Y$.*

A channel $\mathcal{T}_{X\to Y}$ that satisfies the two last conditions will be referred to as a *lambda-majorization channel*.

Furthermore, although the channel $\mathcal{T}$ is not directly a physical channel (it can be, for example, trace-decreasing), it can always be viewed as part of a unital channel $\bar{\mathcal{E}}$, in the sense that $\mathcal{T}$ can be obtained by projection onto specific subspaces and tracing out the ancilla A of the channel $\bar{\mathcal{E}}$ (see Appendix A 2). In turn, unital channels are a (strict [37]) superset of the noisy operations. Recall that our task is to find a lower bound on the work cost of all possible processes allowed in our framework, which we will do by optimizing the work cost over all processes that perform a given state transition.

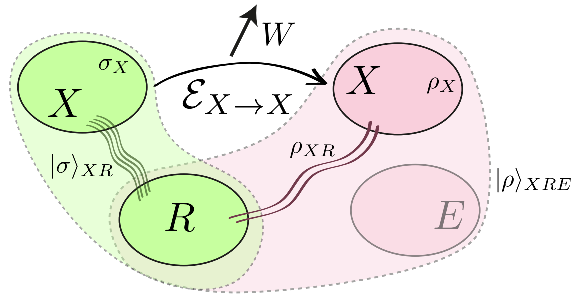

FIG. 3: Our main result gives a fundamental lower bound on the work cost W of a process transforming a state σ_X (purified by a ficiticious $|\sigma\rangle_{XR}$) into a new state ρ_{XR} obtained by applying a process $\mathcal{E}_{X\to X}$. The lower bound to the work cost is given by the entropy that the process $\mathcal{E}$ has to dump into the environment E (in which ρ_{XR} is purified), as measured by the Rényi-zero conditional entropy $H_0(E|X)_\rho$.

However, instead of considering only the unital channels $\bar{\mathcal{E}}$ that are noisy operations, we will relax this last condition and consider all unital channels $\bar{\mathcal{E}}$, and thus allow the optimization to range over all $\mathcal{T}$ that satisfy the conditions of the above proposition. This will make our lower bound even stronger, by showing that the lower bound still holds even if we relax somewhat the assumptions in our framework.

Main Result.—We are now ready to derive our main result. Consider a system X in the state σ_X. This system can always be purified by a reference system, R, in a pure joint state $|\sigma\rangle_{XR}$.

Allowing actions defined by our framework on X, we will study the transition of this state to a state ρ_{XR}, by applying a process $\mathcal{T}_{X\to X}$. The systems are depicted in Figure 3.

The task we would like to solve is the following. Given σ_X and a process $\mathcal{E}_{X\to X}$, and given a purification $|\sigma\rangle_{XR}$ of σ_X and an output state $\rho_{XR} = \mathcal{E}(\sigma_{XR})$, we would like to find the least amount of work W one has to pay for any process in our framework that implements the action of $\mathcal{E}$ on σ. As we have seen in the previous section, we can formulate within our framework all possible processes as lambda-majorizations, so our task is actually to find the best λ such that $\sigma_X \xrightarrow{\lambda} \rho_X$, with the corresponding lambda-majorization channel $\mathcal{T}$ from Prop. 2 satisfying $\mathcal{T}(\sigma_{XR}) = \rho_{XR}$.

Our main result gives an upper bound on the optimal amount of work that can be extracted by this transition, or equivalently, a lower bound on the minimum amount of work that will have to be paid in order to perform the transition. The main result follows directly from following technical proposition.

We are given an input state σ_X and a process $\mathcal{E}_{X\to X}$. Let $|\sigma\rangle_{XR}$ be a purification of σ_X, and let $\rho_{XR} = \mathcal{E}_{X\to X}(\sigma_{XR})$. Let also ρ_{XRE} be a purification of ρ_{XR} in an environment system E. The Rényi-zero entropy

$H_0(E|X)_\rho$ [26, 38] is defined by

$$H_0(E|X)_\rho = \max_{\substack{\omega_X \geqslant 0 \\ \operatorname{tr}\omega_X = 1}} \operatorname{tr}[\Pi_{XE}\,\omega_X] \ , \tag{3}$$

where Π_{XE} is the projector on the support of ρ_{XE}.

Proposition 3. *Then the λ-majorization $\sigma_X \xrightarrow{\lambda} \rho_X$ holds, with the channel $\mathcal{T}_{X\to X}$ from Prop 2 satisfying $\mathcal{T}(\sigma_{XR}) = \rho_{XR}$, if and only if $\lambda \leqslant -H_0(E|X)_\rho$.*

Main Result. *Any process in our framework acting on system X that implements the channel $\mathcal{E}$ when given input σ_X (or equivalently, that brings the state σ_{XR} to the state ρ_{XR}) has to cost at least $kT\ln(2)\cdot H_0(E|X)$ work.*

In other words, *the minimal work cost of a transition from σ to ρ is given by the amount of (information-theoretic) entropy dumped into the environment, conditioned on the output of the computation.* This is precisely the quantitative generalization to correlated quantum systems of the original Landauer's principle [1].

It is worth noting that instead of specifying the channel $\mathcal{E}$, we may also simply specify the output state ρ_{XR}, which completely determines the process (on the support of σ_X) since it is the Choi-Jamiołkowski state corresponding to $\mathcal{E}$ rescaled by σ_X ($\rho_{XR} = \mathcal{E}(\sigma_{XR})$). One can thus understand the input to the problem to actually be a bipartite state ρ_{XR}, such that ρ_X is the required output, ρ_R is the input that will be fed into the process, and any correlations between X and R specify parts of the output that we wish be preserved and not be modified, or thermalized, by the process.

The full proof of Prop. 3 is provided in the appendix. We provide the general idea of the proof in the following.

Proof Sketch of the Main Result. The main idea of the proof is to write the optimization problem as a semidefinite program for the variables $\alpha = 2^{-\lambda}$, $T_{XX'}$ (the Choi-Jamiołkowski representation of $\mathcal{T}_{X\to X'}$), and the dual variables $\omega_{X'}$, X_X and $Z_{X'R}$. Let $(\cdot)^{t_X}$ denote the partial transpose operation on X. The optimal extracted work λ is given by the following semidefinite program:

<u>Primal</u>

`minimize:` α

`subject to:`

$$\operatorname{tr}_X[T_{XX'}] \leqslant \alpha \mathbb{1}_{X'} \quad : \omega_{X'}$$
$$\operatorname{tr}_{X'}[T_{XX'}] \leqslant \mathbb{1}_X \quad : X_X$$
$$\operatorname{tr}_X\left[T_{XX'}\sigma_{XR}^{t_X}\right] = \rho_{X'R} \quad : Z_{X'R}$$

<u>Dual</u>

`maximize:` $\operatorname{tr}(Z_{X'R}\,\rho_{X'R}) - \operatorname{tr} X_X$

`subject to:`

$$\operatorname{tr}\omega_{X'} \leqslant 1$$
$$\operatorname{tr}_R\left[\sigma_{XR}^{t_X} Z_{X'R}\right] \leqslant \mathbb{1}_X \otimes \omega_{X'} + X_X \otimes \mathbb{1}_{X'}$$

The optimal value $\alpha = 2^{H_0(E|X')_\rho}$ is achieved (see Appendix B) by the completely positive map $\mathcal{T}_{X\to X'} = \operatorname{tr}_E\left[V_{X\to X'E}(\cdot)V^\dagger\right]$, where $V_{X\to X'E}$ is the partial isometry with minimal support relating σ_{XR} to $\rho_{X'ER}$ (both being purifications of the same $\sigma_R = \rho_R$).

While it is clear from the formulation of our problem that $\mathcal{T}$ is already completely determined on the support of σ_X (expressed by the condition $\mathcal{T}(\sigma_{XR}) = \rho_{XR}$), the optimization over $\mathcal{T}$ is done in order to (at least formally) find the optimal action on the complement of the support of σ_X. Also, the formulation of a lambda-majorization problem as a semidefinite program is a more general toolbox that could be used in the case where the mapping is not completely determined and where arbitrary additional semidefinite conditions can be imposed at will.[1]

Allowing a Probability of Error. A "smooth" version of the result is straightforward to obtain. In this case, we allow the actual process to not exactly implement $\mathcal{E}$, but only approximate it well. The best strategy to detect this failure is to prepare $|\sigma\rangle_{XR}$ and send σ_X into the process, and then perform a measurement on ρ_{XR}. To ensure the probability of error does not exceed ε, the trace distance between the ideal output of the process ρ_{XR} and the actual output $\hat{\rho}_{XR}$ must not exceed ε. We can apply our main result to the approximate process that brings σ to $\hat{\rho}$, and lower bound the work cost of that process by

$$\begin{aligned} W(\sigma\to\hat{\rho}) &\geqslant H_0(E|X)_{\hat{\rho}}\cdot kT\ln(2) \\ &\geqslant H_{\max}(E|X)_{\hat{\rho}}\cdot kT\ln(2)\ , \end{aligned} \tag{6}$$

where the second inequality is shown in [39] and involves the max entropy measure $H_{\max}$ as defined in [27, 28]. For any $\bar{\varepsilon} \geqslant 0$, the smooth max entropy $H_{\max}^{\bar{\varepsilon}}$ is defined as

$$H_{\max}^{\bar{\varepsilon}}(E|X)_\rho = \min_{\hat{\rho}\overset{\bar{\varepsilon}}{\approx}\rho} \max_{\substack{\tau_X\geqslant 0\\ \operatorname{tr}\tau_X=1}} \log F^2(\hat{\rho}_{EX}, \mathbb{1}_E\otimes\tau_X)\,, \tag{7}$$

where the first optimization ranges over all $\hat{\rho}_{EX}$ such that $F^2(\hat{\rho},\rho) \geqslant 1-\bar{\varepsilon}^2$ and where $F(\rho,\hat{\rho}) = \|\sqrt{\rho}\sqrt{\hat{\rho}}\|_1$ is the fidelity between the quantum states ρ and $\hat{\rho}$ [40]. We write $H_{\max}$ to indicate $H_{\max}^{\bar{\varepsilon}}$ with $\bar{\varepsilon} = 0$.

If we optimize (6) over all possible channels $\mathcal{T}$ that output such $\hat{\rho}_{XR}$, we obtain a bound on the extractable

---

[1] For example, instead of fixing the process with $\mathcal{T}(\sigma_{XR}) = \rho_{XR}$, one may have instead required that $\mathcal{T}(\sigma_X) = \rho_X$ for given σ_X and ρ_X, not specifying and optimizing over what happens to correlations between the input and the output (or, equivalently, one could optimize over ρ_{XR} with fixed reductions ρ_X and ρ_R). In that case, the semidefinite program can be used to obtain bounds to the optimal value. This also implies that the "relative mixedness" introduced in [19] can be formulated as a semidefinite program.

work with a probability of error ε,

$$\begin{aligned} W &\geqslant \min_{\hat{\rho}_{XR} \overset{\varepsilon}{\approx} \rho_{XR}} H_{\max}(E|X)_{\hat{\rho}} \cdot kT\ln(2) \\ &\geqslant \min_{\hat{\rho}_{XRE} \overset{\bar{\varepsilon}}{\approx} \rho_{XRE}} H_{\max}(E|X)_{\hat{\rho}} \cdot kT\ln(2) \\ &= H_{\max}^{\bar{\varepsilon}}(E|X)_{\rho} \cdot kT\ln(2) \ , \end{aligned} \tag{8}$$

where the first optimization ranges over all $\hat{\rho}_{XR}$ such that the trace distance $\frac{1}{2}\|\hat{\rho}_{XR} - \rho_{XR}\|_1 \leqslant \varepsilon$, and where the second optimization ranges over all $\hat{\rho}_{XRE}$ such that $F^2(\rho_{XRE}, \hat{\rho}_{XRE}) \geqslant 1 - \bar{\varepsilon}^2$, with $\bar{\varepsilon} = \sqrt{2\varepsilon}$.

Tightness of the Bound.—The bound given in the main result is tight up to error terms of the order of $\log \frac{1}{\varepsilon}$. Indeed, let's consider the following simple process: one appends a large enough ancilla A_E in a pure state to the input, so that we have our systems in the state $\sigma_{XRA_E} = |0\rangle_{A_E} \otimes |\sigma\rangle_{XR}$. Let us consider a purification $|\rho\rangle_{XRA_E}$ of ρ_{XR}. Since the reduced state on R of both these states are the same, $\sigma_R = \rho_R$, there exists a unitary U acting on $X \otimes A_E$ such that $|\rho\rangle_{XRA_E} = U|\sigma\rangle_{XRA_E}$. So we can apply this unitary onto our input at no work cost, and we are left with $|\rho\rangle_{XRA_E}$ on our systems. We then apply the protocol proposed by del Rio *et al.* [5] on the system A_E, using the system X as a memory we have access to, in order to erase the ancilla A_E back to a pure state. Recall that their process acheives this task without modifying the reduced state ρ_{XR}, and at a work cost $kT\ln(2)H_{\max}^{\varepsilon}(A_E|R) + O\left(\log\frac{1}{\varepsilon}\right)$. It is also straightforward to note that their protocol can be carried out within our framework. Thus, up to error terms of the order of the logarithm of the error probability, our bound given by (8) is tight.

Special Cases.—From our main result we can recover several some special cases of specific interest as corollaries.

Von Neumann Limit. As we have seen in the introduction, considerable previous work has focused on the limit cases where many i.i.d. systems are provided. In such a case, the process $\mathcal{E}^{\otimes n}$ is applied on n independent copies of the input $\sigma^{\otimes n}$, and outputs $\rho^{\otimes n}$. Say we tolerate a probability of error ε. We may simply apply our (smoothed) main result to get an expression for our bound on the work cost,

$$W \geqslant H_{\max}^{\bar{\varepsilon}}(E^n|X^n)_{\rho^{\otimes n}} \cdot kT\ln(2) \ , \tag{9}$$

however it is known that the smooth entropies converge to the von Neumann entropy in the i.i.d. limit [41],

$$\lim_{\bar{\varepsilon}\to 0}\lim_{n\to\infty} \frac{1}{n} H_{\max}^{\bar{\varepsilon}}(E^n|X^n)_{\rho^{\otimes n}} = H(E|X)_{\rho} \ , \tag{10}$$

which allows us to simplify the expression to

$$H(E|X)_{\rho} = H(EX)_{\rho} - H(X)_{\rho} = H(X)_{\sigma} - H(X)_{\rho} \ ,$$

where the last equality holds because ρ_{EX} and σ_X have the same spectrum being both purifications of the same $\rho_R = \sigma_R$. We conclude that in the asymptotic i.i.d. case, the work cost of such a process is simply given by the difference of entropy between the initial and final state,

$$W \geqslant [H(\text{initial state}) - H(\text{final state})]\, kT\ln(2) \ . \tag{11}$$

We emphasize that in this case the exact process is not relevant, and only the input and output states matter. If one considers the example given in the introduction with (a) the identity channel and (b) a replacement map, and apply these processes on n *independent copies* of the distribution described in Figure 1, then in this regime both processes cost no work.

Erasure of a Quantum System Using a Quantum Memory. Consider the setting proposed in [5], where a system S is correlated to a system M in a joint state σ_{SM}, and where our task is to erase S while preserving the reduced state on M and any possible correlations of M with other systems. Formally, given a purification σ_{SMR} of σ_{SM}, we are looking for a process that will bring this state to the state $\rho_{SMR} = |0\rangle\langle 0|_S \otimes \sigma_{MR}$, i.e. we require the process to preserve σ_{MR}. In [5] a process is proposed that performs this task at work cost

$$kT\ln(2)\, H_{\max}^{\varepsilon}(S|M)_{\sigma} + O\left(\log\tfrac{1}{\varepsilon}\right) \ ,$$

where $H_{\max}^{\varepsilon}$ is the smooth max entropy [27–29].

This is a special case of the general case considered above, simply by considering X to be the joint system of S and the memory M, $\mathscr{H}_X = \mathscr{H}_S \otimes \mathscr{H}_M$. Note that we have $\rho_{SMR} = |0\rangle\langle 0|_S \otimes \sigma_{MR}$, purified by $|\rho\rangle_{SMRE} = |0\rangle_S \otimes |\rho\rangle_{MRE}$, where $|\rho\rangle_{MRE} = U_{S\to E}|\sigma\rangle_{SMR}$ and $U_{S\to E}$ is an isometry from S to E.

Then the bound on the work cost, tolerating a probability of error of at most ε, is

$$\begin{aligned} W &\geqslant H_{\max}^{\varepsilon}(E|SM)_{\rho} \cdot kT\ln(2) \\ &= H_{\max}^{\varepsilon}(E|M)_{\rho} \cdot kT\ln(2) \\ &= H_{\max}^{\varepsilon}(S|M)_{\sigma} \cdot kT\ln(2) \ , \end{aligned} \tag{12}$$

where the first equality follows because ρ is pure on S and the second by reversing the isometry U. We can immediately conclude that, within our framework, any process that performs this erasure has to cost at least $kT\ln(2)\, H_{\max}^{\varepsilon}(S|M)_{\sigma}$ work. Thus, the process proposed by del Rio *et al.* is optimal up to logarithmic factors in the error probability ε. Note that if we take the memory M to be trivial i.e. a pure state, then we are in the standard scenario of Landauer erasure on a single system, and we have $W \geq H_{\max}^{\varepsilon}(S)$ which is achievable, recovering the result of [18].

State Transformation while Decoupling from the Reference System. Another special case that we can derive as a corollary is if we consider the process that erases its input and prepares the required output independently. This would occur if we required the output state to be completely uncorrelated to the reference sys-

tem R. Being a replacement map, this process implies that $\rho_{XR} = \rho_X \otimes \rho_R$. In this case, any third party R that would have been correlated to the input is now completely uncorrelated to the output.

Again, we may simply apply our main result with the additional condition that $\rho_{XR} = \rho_X \otimes \rho_R$. In this case, the purification of ρ_{XR}, ρ_{XRE}, takes a special form due to the tensor product structure, with the E system split into two E_R and E_X systems ($E = E_R \otimes E_X$),

$$|\rho\rangle_{XRE} = |\psi\rangle_{XE_X} \otimes |\phi\rangle_{RE_R} \ , \tag{13}$$

where $|\psi\rangle_{XE_X}$ and $|\phi\rangle_{RE_R}$ are purifications of ρ_X and ρ_R, respectively.

The lower bound on the work cost W, given by our main result and tolerating a probability of error of at most ε, then reads

$$W \geqslant H_{\max}^{\bar{\varepsilon}}\left(E|X\right)_\rho = H_{\max}^{\bar{\varepsilon}}\left(E_R\right)_{|\phi\rangle} + H_{\max}^{\bar{\varepsilon}}\left(E_X|X\right)_{|\psi\rangle} ,$$

where $\bar{\varepsilon} = \sqrt{2\varepsilon}$ and $H_{\max}^{\bar{\varepsilon}}$ is again the smooth max entropy. Now, the spectrum of ρ_{E_R} is exactly the same as the spectrum of ρ_R by the Schmidt decomposition of $|\phi\rangle$. This in turn has the same spectrum as σ_X also by the Schmidt decomposition of ρ_{XR} and because $\rho_R = \sigma_R$. It follows that $H_{\max}^{\bar{\varepsilon}}(E_R)_\rho = H_{\max}^{\bar{\varepsilon}}(X)_\sigma$. Also, by duality of smooth min and max entropies [27], we have $H_{\max}^{\bar{\varepsilon}}\left(E_X|X\right)_{|\psi\rangle} = -H_{\min}^{\bar{\varepsilon}}\left(E_X\right)_\rho = -H_{\min}^{\bar{\varepsilon}}\left(X\right)_\rho$, where $H_{\min}^{\bar{\varepsilon}}$ is the smooth min entropy with purified distance smoothing as defined in Ref. [28]. In consequence,

$$W \geqslant H_{\max}^{\bar{\varepsilon}}\left(X\right)_\sigma - H_{\min}^{\bar{\varepsilon}}\left(X\right)_\rho \ . \tag{14}$$

That is, to transform a state σ to ρ while maximally decoupling ρ from the reference system, then one has to erase σ to a pure state (at cost $H_{\max}^{\bar{\varepsilon}}\left(X\right)_\sigma$), and then prepare ρ (extracting work $H_{\min}^{\bar{\varepsilon}}\left(X\right)_\rho$).

Example: Erasing Part of the W State.—To illustrate some points mentioned above, consider the W state on a system S, a memory M and a reference system R given by

$$|W\rangle_{SMR} = \frac{1}{\sqrt{3}}\left[|001\rangle + |010\rangle + |100\rangle\right]_{SMR} \ . \tag{15}$$

The reduced states on SM and M are respectively given by $\sigma_{SM} = \frac{1}{3}|00\rangle\langle 00| + \frac{2}{3}|\Psi^+\rangle\langle\Psi^+|$ and $\sigma_M = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1|$, where $|\Psi^+\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle + |10\rangle\right)$. By symmetry of the W state, the reduced state on any two or one qubit(s) have the same form.

By actions on S and M, we would like to erase S, leading to the final state on S and M given by $\rho_{SM} = |0\rangle\langle 0| \otimes \sigma_M$. Let us consider two processes that achieve this goal: the first one will preserve correlations with R but will cost work, the second will not cost work but will modify those correlations.

We may directly apply the special case above concerning the erasure of a system conditioned on a memory: the fundamental work cost of such an erasure, if one preserves correlations with a reference system R, is given by $H_0\left(S|M\right)_\sigma$. One may explicitely calculate (see Appendix C) in this case $H_0\left(S|M\right) = \log\frac{2}{3} \approx 0.59$ and thus this process must cost at least this amount of work.

However, one may easily notice that both σ_{SM} and σ_M have the same spectrum $\{2/3, 1/3\}$. This means that there exists a unitary U that performs the erasure simply as $|0\rangle\langle 0| \otimes \sigma_M = U\sigma_{SM}U^\dagger$, and this unitary process does not cost any work. However, the correlations with R are not preserved. Indeed, the unitary sends $|00\rangle$ to $|01\rangle$ and $|\Psi^+\rangle$ to $|00\rangle$, so one explicitely calculates that the state after the process is given by $\rho_{SMR} = U\sigma_{SMR}U^\dagger = \frac{1}{\sqrt{3}}\left[|011\rangle + \sqrt{2}|000\rangle\right] = |0\rangle \otimes \frac{1}{\sqrt{3}}\left[|11\rangle + \sqrt{2}|00\rangle\right]$. We notice that the reduced state on M and R is now pure and differs from initial one, given by $\sigma_{MR} = \frac{1}{3}|00\rangle\langle 00| + \frac{2}{3}|\Psi^+\rangle\langle\Psi^+|$.

Conclusion.—The last few years have seen enormous technological progress in micro- and nano-fabrication, making it possible to construct engines and thermodevices on a microscopic scale [42–50]. In this regime, standard thermodynamic considerations (devised originally for macroscopic devices such as steam engines) are not necessarily applicable. At the same time, with the miniaturization of computing circuits, thermodynamic aspects of information processing have become increasingly relevant. In fact, the heat dissipated by processors is one of the main barriers limiting their performance. Along with these developments, researchers have started to investigate the laws of thermodynamics from an information-theoretic perspective [51–57].

The present work adds to this line of research, providing a rigorous quantitative relationship between information theory and thermodynamics. One of our main findings is that this relationship is more intricate than what previous results (which focused on averaged quantities) may have suggested. In particular, it turns out that the thermodynamic cost of a given information-processing task not only depends on the input and output state, but also on the correlation between them. While this correlation-dependence disappears in certain asymptotic limits, it cannot be neglected in general and, in fact, may become arbitrarily large.

Acknowledgements.—The authors would like to thank Johan Åberg, Lídia del Rio and Joe Renes for many enlightening discussions. PhF, FD and RR were supported by the Swiss National Science Foundation (SNSF) through the National Centre of Competence in Research "Quantum Science and Technology" and through grant No. 200020-135048, and by the European Research Council through grant No. 258932. FD was also supported by the SNSF through grants PP00P2-128455 and 20CH21-138799, as well as by the German Science Foundation (grant CH 843/2-1). JO is funded by the Royal Society of London.

APPENDIX

Appendix A: Formal Approach to Lambda-Majorization

1. Preliminaries and Main Definition

Let $\mathscr{H}_X$, $\mathscr{H}_Y$ be two subspaces of a finite-dimensional Hilbert space $\mathscr{H}_Z$, and let $\mathscr{H}_A$, $\mathscr{H}_B$ be two subspaces of a finite-dimensional Hilbert space $\mathscr{H}_C$. Let $d_{(\cdot)}$ denote the dimensions of the various Hilbert spaces $\mathscr{H}_{(\cdot)}$ and specifically let $d = d_Z = \dim \mathscr{H}_Z$. Denote by $\mathscr{L}(\mathscr{H})$ the set of linear hermitian operators on $\mathscr{H}$, by $\mathscr{P}(\mathscr{H})$ the set of positive semidefinite operators on $\mathscr{H}$, and by $\mathscr{S}_=(\mathscr{H})$ those operators in $\mathscr{P}(\mathscr{H})$ that have unit trace. Let also $\lambda_i(\rho)$ denote the i-th eigenvalue of ρ (in no particular order), and $\lambda_i^{\downarrow}(\rho)$ denote the i-th eigenvalue of ρ taken in decreasing order.

Majorization is discussed in detail in Refs. [31, 32, 58].

Majorization. *A matrix $\sigma \in \mathscr{P}(\mathscr{H}_Z)$ is said to* majorize *$\rho \in \mathscr{P}(\mathscr{H}_Z)$, denoted by $\sigma \succ \rho$, if for all k, $\sum_{i=1}^k \lambda_i^{\downarrow}(\sigma) \geqslant \sum_{i=1}^k \lambda_i^{\downarrow}(\rho)$, and if $\operatorname{tr}\sigma = \operatorname{tr}\rho$.*

The notion of majorization defines a (partial) order relation on $\mathscr{P}(\mathscr{H}_Z)$. When considering the set of density matrices $\mathscr{S}_=(\mathscr{H}_Z)$, there is a "least" element: the fully mixed state, $\frac{1}{d}\mathbb{1}_Z$.

Weak Submajorization. *A matrix $\sigma \in \mathscr{P}(\mathscr{H}_Z)$ is said to* weakly submajorize *$\rho \in \mathscr{P}(\mathscr{H}_Z)$, denoted by $\sigma \succ_w \rho$, if for all k, $\sum_{i=1}^k \lambda_i^{\downarrow}(\sigma) \geqslant \sum_{i=1}^k \lambda_i^{\downarrow}(\rho)$.*

Remark that if $\sigma, \rho \in \mathscr{S}_=(\mathscr{H}_Z)$, then the concept of weak submajorization is equivalent to regular majorization simply because the traces of these matrices are already equal to unity.

Doubly Stochastic Matrix. *A $d \times d$ matrix S is* doubly stochastic *if ${S_i}^j \geqslant 0$, $\sum_i {S_i}^j = 1\ \forall j$ and $\sum_j {S_i}^j = 1\ \forall i$.*

Doubly Substochastic Matrix. *A $n \times m$ matrix B is* doubly substochastic *if ${B_i}^j \geqslant 0$, $\sum_i {B_i}^j \leqslant 1\ \forall j$ and $\sum_j {B_i}^j \leqslant 1\ \forall i$.*

The following theorem is due to Hardy, Littlewood and Pólya [59].

Theorem 4 (Hardy, Littlewood, and Pólya, 1929)**.** *Let $\sigma, \rho \in \mathscr{P}(\mathscr{H}_Z)$. Then $\sigma \succ \rho$ if and only if there exists a $d \times d$ doubly stochastic matrix ${S_i}^j$ such that $\lambda_i(\rho) = \sum_j {S_i}^j \lambda_j(\sigma)$.*

A similar theorem is obtained for weak submajorization and doubly substochastic matrices [31].

Proposition 5. *Let $\sigma \in \mathscr{P}(\mathscr{H}_X)$ and $\rho \in \mathscr{P}(\mathscr{H}_Y)$. Then $\sigma \succ_w \rho$ if and only if there exists a $d_Y \times d_X$ doubly substochastic matrix ${B_i}^j$ such that $\lambda_i(\rho) = \sum_j {B_i}^j \lambda_j(\sigma)$.*

Majorization defines a partial order on states and has a "smallest" element, the fully mixed state. Also, a pure state majorizes any other state.

Proposition 6. *Majorization is preserved by direct sums and tensor products, i.e. if $\sigma \succ \rho$ and $\sigma' \succ \rho'$, then $\sigma \oplus \sigma' \succ \rho \oplus \rho'$ and $\sigma \otimes \sigma' \succ \rho \otimes \rho'$. The same holds for weak submajorization.*

A proof for the direct sum of two vectors can be found in [31, Cor. II.1.4]. We provide here an alternative proof along with the tensor product case.

Proof. Let S_i^j and $S_i'^j$ be doubly stochastic matrices such that $\lambda_i(\rho) = \sum_j S_i^j \lambda_j(\sigma)$ and $\lambda_i(\rho') = \sum_j S_i'^j \lambda_j(\sigma')$. Then $S \oplus S'$ is also doubly stochastic and satisfies $\lambda_i(\rho \oplus \rho') = \sum_j (S \oplus S')_i^j \lambda_j(\sigma \oplus \sigma')$, because the vectors of eigenvalues of the direct sum are simply the direct sums of the individual vector of eigenvalues. This shows that $\sigma \oplus \sigma' \succ \rho \oplus \rho'$.

Analogously, $S \otimes S'$ satisfies $\lambda_{ii'}(\rho \otimes \rho') = \lambda_i(\rho)\lambda_{i'}(\rho') = \sum_{jj'} S_i^j \lambda_j(\sigma)\, S_{i'}^{j'} \lambda_{j'}(\sigma') = \sum_{jj'} (S \otimes S')_{ii'}^{jj'}\, \lambda_{jj'}(\sigma \otimes \sigma')$. $S \otimes S'$ is doubly stochastic, $\sum_{ii'} (S \otimes S')_{ii'}^{jj'} = \sum_{ii'} S_i^j S_{i'}^{j'} = 1$ and $\sum_{jj'} (S \otimes S')_{ii'}^{jj'} = \sum_{jj'} S_i^j S_{i'}^{j'} = 1$.

The same proof holds for doubly substochastic matrices, so majorization may be replaced by weak submajorization in the proposition. □

We are now all set for a formal definition of lambda-majorization.

Let $\lambda \in \mathbb{R}$ and let $\lambda_1, \lambda_2 \geqslant 0$ such that $\lambda = \lambda_1 - \lambda_2$ and 2^{λ_1}, 2^{λ_2} are integers. (The case when 2^{λ} is irrational will be discussed later.) Take $\mathscr{H}_C$ of size greater than both 2^{λ_1} and 2^{λ_2} and let $\mathscr{H}_A$ and $\mathscr{H}_B$ be subspaces of $\mathscr{H}_C$ of respective dimensions 2^{λ_1} and 2^{λ_2}.

Lambda-Majorization. *For $\sigma \in \mathscr{P}(\mathscr{H}_X)$ and $\rho \in \mathscr{P}(\mathscr{H}_Y)$, we say that σ* λ-majorizes *ρ, denoted by $\sigma \xrightarrow{\lambda} \rho$, if there exists such λ_1, λ_2 such that $2^{-\lambda_1}\mathbb{1}_A \otimes \sigma \succ_w 2^{-\lambda_2}\mathbb{1}_B \otimes \rho$. Here $\mathbb{1}_A$, $\mathbb{1}_B$ are the projectors onto the respective subspaces $\mathscr{H}_A$ and $\mathscr{H}_B$ embedded in $\mathscr{H}_C$, of respective dimensions 2^{λ_1}, 2^{λ_2}. Likewise, σ and ρ are considered as living in $\mathscr{H}_Z$ by padding them with zero eigenvalues as necessary.*

We have assumed here that 2^{λ} is rational. If 2^{λ} is irrational, we say that σ λ-majorizes ρ if for all rational $2^{\lambda'}$ with $\lambda' < \lambda$, then $\sigma \xrightarrow{\lambda'} \rho$.

The following proposition guarantees that the definition above does not depend on the exact values of λ_1 and λ_2 but only on their difference. This is the same as saying that a fully mixed state cannot act as a catalyst.

Proposition 7. *For any $\sigma, \rho \in \mathscr{P}(\mathscr{H}_Z)$, and for any n, we have $\sigma \succ_w \rho$ if and only if $\sigma \otimes \frac{\mathbb{1}_n}{n} \succ_w \rho \otimes \frac{\mathbb{1}_n}{n}$.*

Proof. If $\sigma \succ_w \rho$, then the majorization passes over the tensor product, and thus proves the claim. Conversely, if $\sigma \otimes \frac{\mathbb{1}_n}{n} \succ_w \rho \otimes \frac{\mathbb{1}_n}{n}$, then in particular, for any $k \leqslant d$,

$$\sum_{i=1}^{n\cdot k} \lambda_i^{\downarrow}(\tfrac{\mathbb{1}_n}{n} \otimes \sigma) \geqslant \sum_{i=1}^{n\cdot k} \lambda_i^{\downarrow}(\tfrac{\mathbb{1}_n}{n} \otimes \rho) \ . \tag{A1}$$

(d is the maximum rank of σ or ρ.) But $\lambda_{in}^{\downarrow}(\frac{\mathbb{1}_n}{n} \otimes \sigma) = \frac{1}{n}\lambda_i^{\downarrow}(\sigma)$ and thus

$$\sum_{i=1}^{k} \lambda_i^{\downarrow}(\sigma) \geqslant \sum_{i=1}^{k} \lambda_i^{\downarrow}(\rho) \ . \qquad \square$$

The following proposition is a direct consequence of the definition of lambda-majorization, and just states that you can move around randomness into or out of the ancillas in the definition of lambda-majorization.

Proposition 8. *For any $\sigma \in \mathscr{P}(\mathscr{H}_X)$, $\rho \in \mathscr{P}(\mathscr{H}_Y)$, and for any $\lambda \in \mathbb{R}$, $n > 0$, we have*

$$\tfrac{1}{n}\mathbb{1}_n \otimes \sigma \xrightarrow{\lambda} \rho \ \Leftrightarrow\ \sigma \xrightarrow{\lambda + \log n} \rho$$

and

$$\sigma \xrightarrow{\lambda - \log n} \rho \ \Leftrightarrow\ \sigma \xrightarrow{\lambda} \tfrac{1}{n}\mathbb{1}_n \otimes \rho \ .$$

Similarly to Thm. 4 and to Prop. 5, it is possible to characterize lambda-majorization by the existence of a matrix relating the vector of eigenvalues that satisfies some specific normalization conditions.

Proposition 9. *Let $\sigma \in \mathscr{P}(\mathscr{H}_X)$ and $\rho \in \mathscr{P}(\mathscr{H}_Y)$. Then $\sigma \xrightarrow{\lambda} \rho$ if and only if there exists a $d_Y \times d_X$ matrix ${T_i}^k$ such that $\lambda_i(\rho) = \sum_k {T_i}^k \lambda_k(\sigma)$, satisfying ${T_i}^k \geqslant 0$, $\sum_i {T_i}^k \leqslant 1$, and $\sum_k {T_i}^k \leqslant 2^{-\lambda}$.*

Proof of Prop. 9. Suppose $2^{-\lambda_1}\mathbb{1}_A \otimes \sigma \succ_w 2^{-\lambda_2}\mathbb{1}_B \otimes \rho$ with $\lambda = \lambda_1 - \lambda_2$. Then there exists a doubly substochastic matrix S_{bi}^{ak} such that

$$\lambda_{bi}\big(2^{-\lambda_2}\mathbb{1}_B \otimes \rho\big) = \sum_{ak} S_{bi}^{ak}\, \lambda_{ak}\big(2^{-\lambda_1}\mathbb{1}_A \otimes \sigma\big) \ ,$$

with $S_{bi}^{ak} \geqslant 0$, $\sum_{bi} S_{bi}^{ak} \leqslant 1$ and $\sum_{ak} S_{bi}^{ak} \leqslant 1$. (Indices a and b refer to the mixed ancillas of respective sizes 2^{λ_1} and 2^{λ_2}. Since we are considering weak submajorization, we can safely ignore all zero eigenvalues and consider only the subspaces (of different sizes on the left and right hand side of the majorization) on which σ, ρ, $\mathbb{1}_A$ and $\mathbb{1}_B$ have support, as in Prop. 5.)

Now we have

$$\begin{aligned} \lambda_i(\rho) &= \sum_b \lambda_{bi}\big(2^{-\lambda_2}\mathbb{1}_B \otimes \rho\big) \\ &= \sum_{a\,b\,k} S_{bi}^{ak}\, \lambda_{ak}\big(2^{-\lambda_1}\mathbb{1}_A \otimes \sigma\big) \\ &= \sum_k \left(\sum_{ab} 2^{-\lambda_1} S_{bi}^{ak}\right) \lambda_k(\sigma) \ , \end{aligned}$$

so one can define

$${T_i}^k = \sum_{ab} 2^{-\lambda_1} S_{bi}^{ak} \ ,$$

which fulfills $\lambda_i(\rho) = \sum_k {T_i}^k \lambda_k(\sigma)$. Because S is doubly substochastic, and using the fact that indices a (resp. b) range to 2^{λ_1} (2^{λ_2}), the matrix T satisfies

$$\sum_i {T_i}^k = \sum_{i\,a\,b} 2^{-\lambda_1} S_{bi}^{ak} = \sum_a 2^{-\lambda_1} \sum_{bi} S_{bi}^{ak} \leqslant 1 \ ,$$

as well as

$$\begin{aligned} \sum_k {T_i}^k = \sum_{k\,a\,b} 2^{-\lambda_1} S_{bi}^{ak} &= \sum_b 2^{-\lambda_1} \sum_{ak} S_{bi}^{ak} \\ &\leqslant \sum_b 2^{-\lambda_1} = 2^{-\lambda} \ . \end{aligned}$$

Additionally, ${T_i}^k \geqslant 0$ because $S_{bi}^{ak} \geqslant 0$.

Conversely, suppose that a matrix ${T_i}^k$ exists, with ${T_i}^k \geqslant 0$, $\sum_i {T_i}^k \leqslant 1$, $\sum_k {T_i}^k \leqslant 2^{-\lambda}$, and $\lambda_i(\rho) = \sum_k {T_i}^k \lambda_k(\sigma)$. Let λ_1, λ_2 such that $\lambda = \lambda_1 - \lambda_2$ and such that 2^{λ_1}, 2^{λ_2} are integers. Then let $S_{bi}^{ak} = 2^{-\lambda_2} {T_i}^k$ for all a, b. Then $S_{bi}^{ak} \geqslant 0$ and S satisfies

$$\sum_{ak} S_{bi}^{ak} = 2^{-\lambda_2} \sum_{ak} {T_i}^k \leqslant 2^{-\lambda_2}\Big(\sum_a 1\Big)\, 2^{-\lambda} = 1 \ ,$$

as well as

$$\sum_{bi} S_{bi}^{ak} = 2^{-\lambda_2} \sum_{bi} {T_i}^k \leqslant 2^{-\lambda_2}\Big(\sum_b 1\Big) = 1 \ .$$

The required weak submajorization for the desired lambda-majorization is provided by this doubly substochastic matrix,

$$\begin{aligned} \lambda_{bi}\big(2^{-\lambda_2}\mathbb{1}_B \otimes \rho\big) &= 2^{-\lambda_2}\lambda_i(\rho) = 2^{-\lambda_2} \sum_k {T_i}^k \lambda_k(\sigma) \\ &= 2^{-\lambda_2} \sum_k {T_i}^k \sum_a \lambda_{ak}\big(2^{-\lambda_1}\mathbb{1}_A \otimes \sigma\big) \\ &= \sum_{ak} S_{bi}^{ak} \lambda_{ak}\big(2^{-\lambda_1}\mathbb{1}_A \otimes \sigma\big) \ . \qquad \square \end{aligned}$$

2. Formulation of Lambda-Majorization in Terms of Channels

Majorization can also be characterized in terms of unital, trace-preserving completely positive maps [33–36].

Proposition 10. *Two positive semidefinite matrices σ and ρ satisfy $\sigma \succ \rho$ if and only if there exists a trace-preserving, unital, completly positive map $\mathcal{E}$ satisfying $\mathcal{E}(\sigma) = \rho$.*

Similarly, one can prove an analogous characterization of weak submajorization. The proof of this proposition will be given later.

Proposition 11. *Let $\sigma \in \mathscr{P}(\mathscr{H}_X)$ and $\rho \in \mathscr{P}(\mathscr{H}_Y)$. Then $\sigma \succ_w \rho$ if and only if there exists a completely positive map $\mathcal{E}_{X\to Y} : \mathscr{L}(\mathscr{H}_X) \to \mathscr{L}(\mathscr{H}_Y)$ such that $\mathcal{E}_{X\to Y}(\sigma) = \rho$, with $\mathcal{E}$ satisfying $\mathcal{E}_{X\to Y}(\mathbb{1}_X) \leqslant \mathbb{1}_Y$ and $\mathcal{E}^\dagger_{X\to Y}(\mathbb{1}_Y) \leqslant \mathbb{1}_X$.*

Let's say that $\mathcal{E}_{X\to Y}$ is *subunital* if $\mathcal{E}_{X\to Y}(\mathbb{1}_X) \leqslant \mathbb{1}_Y$. Then the two conditions on the structure of the channel $\mathcal{E}_{X\to Y}$ in the above proposition require the channel to be subunital and trace-nonincreasing.

A subunital trace-nonincreasing completely positive map can always be seen as part of a unital, trace-preserving completely positive map on a larger Hilbert space. This is analogous of the result that doubly substochastic matrices are submatrices of stochastic matrices [31].

Proposition 12. *Let $\mathcal{E}_{Z\to Z}$ be a unital, trace-preserving completely positive map. Let $\mathscr{H}_X$ and $\mathscr{H}_Y$ be two subspaces of $\mathscr{H}_Z$ and let $\mathbb{1}_X$ and $\mathbb{1}_Y$ be the projector onto those spaces, respectively. Then the channel $\mathcal{E}'_{X\to Y}(\cdot) = \mathbb{1}_Y \mathcal{E}(\mathbb{1}_X \cdot \mathbb{1}_X)\mathbb{1}_Y$ is subunital and trace-decreasing.*

Conversely, let $\mathcal{E}'_{X\to Y}$ be any trace-decreasing, subunital completely positive map. Let $\mathscr{H}_Z = \mathscr{H}_X \oplus \mathscr{H}_Y$, $G_Y = \mathbb{1}_Y - \mathcal{E}'_{X\to Y}(\mathbb{1}_X) \geqslant 0$, and $H_X = \mathbb{1}_X - \mathcal{E}'^\dagger(\mathbb{1}_Y) \geqslant 0$. Then the channel defined by

$$\begin{aligned}
&\mathcal{E}_{Z\to Z}(\cdot)\\
&\quad = 0_X \oplus \mathcal{E}'_{X\to Y}(\mathbb{1}_X (\cdot) \mathbb{1}_X)\\
&\quad + \mathcal{E}'^\dagger(\mathbb{1}_Y (\cdot) \mathbb{1}_Y) \oplus 0_Y\\
&\quad + \left(0_X \oplus \sqrt{G_Y}\right)(\cdot)\left(0_X \oplus \sqrt{G_Y}\right)\\
&\quad + \left(\sqrt{H_X} \oplus 0_Y\right)(\cdot)\left(\sqrt{H_X} \oplus 0_Y\right)
\end{aligned}$$

is unital and trace-preserving, and $\mathcal{E}'_{X\to Y}(\cdot) = \mathbb{1}_Y \mathcal{E}(\mathbb{1}_X (\cdot) \mathbb{1}_X)\mathbb{1}_Y$.

In order to generalize this concept to our lambda-majorization, let's introduce the concept of an α-subunital map. These generalize the notion of subunital maps to arbitrary normalizations.

α-subunital Maps. We'll call a map $\mathcal{T}_{X\to Y}$ *α-subunital* if it satisfies $\mathcal{T}_{X\to Y}(\mathbb{1}_X) \leqslant \alpha \mathbb{1}_Y$.

Proposition 13 (Composition of α-subunital maps)**.** *Let $\mathscr{H}_W \in \mathscr{H}_Z$ be another subspace of $\mathscr{H}_Z$ in addition to $\mathscr{H}_X$ and $\mathscr{H}_Y$, and let $\mathcal{T}_{X\to Y}$, $\mathcal{T}'_{Y\to W}$ be trace-nonincreasing maps. Assume that $\mathcal{T}_{X\to Y}$ is α-subunital and that $\mathcal{T}'_{Y\to W}$ is β-subunital. Then their composition $[\mathcal{T}' \circ \mathcal{T}]_{X\to W}$ is $\alpha \cdot \beta$-subunital.*

Proof of Prop. 13. The composition of $\mathcal{T}_{X\to Y}$ and $\mathcal{T}'_{Y\to W}$ is trace-nonincreasing,

$$\mathcal{T}^\dagger\left(\mathcal{T}'^\dagger(\mathbb{1}_W)\right) \leqslant \mathcal{T}^\dagger(\mathbb{1}_Y) \leqslant \mathbb{1}_X \ .$$

Their composition is also $\alpha \cdot \beta$-subunital,

$$\mathcal{T}'_{Y\to W}(\mathcal{T}_{X\to Y}(\mathbb{1}_X)) \leqslant \mathcal{T}'_{Y\to W}(\alpha\, \mathbb{1}_Y) \leqslant \alpha\beta\, \mathbb{1}_W \ . \qquad \square$$

We will now give proofs for Props. 11 and 12, which rely on the following lemma.

Lemma 14. *Let $\mathcal{T}_{Z\to Z}$ be a trace-nonincreasing map that is $2^{-\lambda}$-subunital. Denote by $\mathbb{1}_X$ (resp. $\mathbb{1}_Y$) the projectors onto the subspaces $\mathscr{H}_X$ (resp. $\mathscr{H}_Y$) of $\mathscr{H}_Z$. Then $\mathcal{T}_{X\to Y}$, defined by $\mathcal{T}_{X\to Y}(\cdot) = \mathbb{1}_Y \mathcal{T}_{Z\to Z}(\mathbb{1}_X (\cdot) \mathbb{1}_X)\mathbb{1}_Y$, is also a trace-nonincreasing $2^{-\lambda}$-subunital map.*

Proof of Lemma 14. It suffices to note that the projection map: $(\cdot) \to \mathbb{1}_X (\cdot) \mathbb{1}_X$ (resp. $(\cdot) \to \mathbb{1}_Y (\cdot) \mathbb{1}_Y$) is trace-nonincreasing and subunital. Then apply Prop. 13 twice. $\square$

Proof of Prop. 12. The first part of the proposition follows from the lemma. To prove the converse, let $\mathcal{E}_{Z\to Z}$ as in the proposition text, and notice first that the channel is its own adjoint:

$$\begin{aligned}
\mathcal{E}^\dagger(\cdot) &= \mathcal{E}'^\dagger(\mathbb{1}_Y (\cdot) \mathbb{1}_Y) \oplus 0_Y + 0_X \oplus \mathcal{E}'(\mathbb{1}_X (\cdot) \mathbb{1}_X)\\
&\quad + \left(0_X \oplus \sqrt{G_Y}\right)(\cdot)\left(0_X \oplus \sqrt{G_Y}\right)\\
&\quad + \left(\sqrt{H_X} \oplus 0_Y\right)(\cdot)\left(\sqrt{H_X} \oplus 0_Y\right)\\
&= \mathcal{E}_{X\to Y}(\cdot) \ .
\end{aligned} \tag{A2}$$

The map is unital:

$$\begin{aligned}
\mathcal{E}_{Z\to Z}(\mathbb{1}_Z) = 0_X \oplus (\mathbb{1}_Y - G_Y) + (\mathbb{1}_X - H_X) \oplus 0_Y&\\
+ 0_X \oplus G + H_X \oplus 0_Y = \mathbb{1}_Z&\ ,
\end{aligned}$$

and it is thus trace-preserving because of (A2). The last condition, $\mathcal{E}'_{X\to Y}(\cdot) = \mathbb{1}_Y \mathcal{E}_{Z\to Z}(\mathbb{1}_X (\cdot) \mathbb{1}_X)\mathbb{1}_Y$ is obvious from the definition of $\mathcal{E}_{Z\to Z}$. $\square$

Proof of Prop. 11. By the weak submajorization condition, if $\operatorname{tr}\rho \neq \operatorname{tr}\sigma$, we must have $\operatorname{tr}\rho < \operatorname{tr}\sigma$. Consider an extension space $\mathscr{H}_{Y'} \in \mathscr{H}_Z$ (consider a larger $\mathscr{H}_Z$ if necessary) in which we extend ρ by many small eigenvalues such that $\operatorname{tr}\rho_{Y\oplus Y'} = \operatorname{tr}\sigma$, while still having $\sigma \succ_w \rho_{Y\oplus Y'}$. Now we have a (regular) majorization, $\sigma \succ \rho_{Y\oplus Y'}$, and can apply Prop. 10.

The obtained map, $\mathcal{E}_{Z\to Z}$, is then unital and trace-preserving. It can be restricted by projecting the input

onto $\mathscr{H}_X$ and the output onto $\mathscr{H}_Y$,

$$\mathcal{E}_{X\to Y}(\cdot) = \mathbb{1}_Y\, \mathcal{E}_{Z\to Z}\left(\mathbb{1}_X\,(\cdot)\,\mathbb{1}_X\right)\,\mathbb{1}_Y\ .$$

This restricted operator, by the lemma, is a valid trace-nonincreasing subunital map (take $\lambda = 0$).

Conversely, if $\mathcal{E}_{X\to Y}$ is a subunital trace-nonincreasing completely positive map with $\mathcal{E}_{X\to Y}(\sigma) = \rho$, then one can dilate it with Proposition 12 to a unital, trace-preserving completely positive map $\mathcal{E}_{Z\to Z}$ such that $\mathbb{1}_Y\,\mathcal{E}_{Z\to Z}\,(\sigma\oplus 0_Y)\,\mathbb{1}_Y = \rho$. Note also that the map $(\cdot)\mapsto \mathbb{1}_Y\,(\cdot)\,\mathbb{1}_Y + \mathbb{1}_X\,(\cdot)\,\mathbb{1}_X$ is a pinching [31, p. 50, Prob. II.5.5], so we have $\sigma\oplus 0_Y \succ \mathcal{E}_{Z\to Z}\,(\sigma\oplus 0_Y) \succ \mathbb{1}_X\mathcal{E}_{Z\to Z}\,(\sigma\oplus 0_Y)\,\mathbb{1}_X + \mathbb{1}_Y\mathcal{E}_{Z\to Z}\,(\sigma\oplus 0_Y)\,\mathbb{1}_Y \succ_w \mathbb{1}_X\mathcal{E}_{Z\to Z}\,(\sigma\oplus 0_Y)\,\mathbb{1}_X = \rho$. The last weak submajorization is because some eigenvalues were left out. □

In the same way as lambda majorization can be characterized with differently normalized doubly substochastic maps, it can also be characterized in terms of a differently normalized subunital channel.

Proposition 15. *Let $\sigma\in\mathscr{P}(\mathscr{H}_X)$, $\rho\in\mathscr{P}(\mathscr{H}_Y)$ and $\lambda\in\mathbb{R}$. Then $\sigma \xrightarrow{\lambda} \rho$ if and only if there exists a completely positive map $\mathcal{T}_{X\to Y} : \mathscr{L}(\mathscr{H}_X)\to\mathscr{L}(\mathscr{H}_Y)$ such that $\mathcal{T}_{X\to Y}(\sigma)=\rho$, that is $2^{-\lambda}$-subunital and trace-nonincreasing.*

Proof of Prop. 15. **"⇒".** Assume first that $2^{-\lambda_1}\mathbb{1}_A\otimes\sigma \succ_w 2^{-\lambda_2}\mathbb{1}_B\otimes\rho$, with $\mathscr{H}_A$, $\mathscr{H}_B$ (of respective sizes 2^{λ_1} and 2^{λ_2}) being subsystems of an ancilla system $\mathscr{H}_C$, with $\lambda = \lambda_1 - \lambda_2$.

By Prop. 11, there exists a subunital trace-nonincreasing completely positive map $\mathcal{E}_{AX\to BY}$, such that

$$\mathcal{E}_{AX\to BY}(2^{-\lambda_1}\mathbb{1}_A\otimes\sigma) = 2^{-\lambda_2}\mathbb{1}_B\otimes\rho\ . \tag{A3}$$

Now let the map $\mathcal{T}$ be defined by

$$\mathcal{T}_{X\to Y}\,(\cdot) = \mathrm{tr}_B\left[\mathcal{E}_{AX\to BY}\left(2^{-\lambda_1}\mathbb{1}_A\otimes(\cdot)\right)\right]\ . \tag{A4}$$

This map is trace-nonincreasing,

$$\begin{aligned}\mathcal{T}^\dagger_{X\leftarrow Y}\,(\mathbb{1}_Y) &= 2^{-\lambda_1}\,\mathrm{tr}_A\left[\mathcal{E}^\dagger_{AX\leftarrow BY}\,(\mathbb{1}_{BY})\right]\\ &\leqslant 2^{-\lambda_1}\,\mathrm{tr}_A\,(\mathbb{1}_{AX}) = \mathbb{1}_X\ ,\end{aligned}$$

and $2^{-\lambda}$-subunital,

$$\begin{aligned}\mathcal{T}_{X\to Y}\,(\mathbb{1}_X) = 2^{-\lambda_1}\,\mathrm{tr}_B\left[\mathcal{E}\,(\mathbb{1}_{AX})\right] &\leqslant 2^{-\lambda_1}\,\mathrm{tr}_B\,\mathbb{1}_{BY}\\ &= 2^{-\lambda}\mathbb{1}_Y\ .\end{aligned}$$

The map $\mathcal{T}$ brings σ to ρ,

$$\begin{aligned}\mathcal{T}_{X\to Y}\,(\sigma_X) &= \mathrm{tr}_B\left[\mathcal{E}\left(2^{-\lambda_1}\mathbb{1}_A\otimes\sigma_X\right)\right]\\ &= \mathrm{tr}_B\left(2^{-\lambda_2}\mathbb{1}_B\otimes\rho_Y\right) = \rho_Y\ ,\end{aligned}$$

so that $\mathcal{T}$ satisfies all the claimed properties.

"⇐". To prove the converse, assume that a trace-nonincreasing, $2^{-\lambda}$-subunital map $\mathcal{T}_{X\to Y}$ exists, such that $\mathcal{T}_{X\to Y}(\sigma)=\rho$.

Choose λ_1, λ_2 such that $\lambda=\lambda_1-\lambda_2$ and such that 2^{λ_1}, 2^{λ_2}, are integers. (Again, in case 2^λ is irrational, approximate 2^λ arbitrarily well by rational numbers $2^{\lambda'}$.) Choose $\mathscr{H}_C$ large enough to contain two subspaces $\mathscr{H}_A$ and $\mathscr{H}_B$ of respective dimensions 2^{λ_1} and 2^{λ_2}. Let

$$\mathcal{E}_{AX\to BY}\,(\cdot) = 2^{-\lambda_2}\mathbb{1}_B\otimes\mathcal{T}_{X\to Y}\,(\mathrm{tr}_A\,(\cdot))\ . \tag{A5}$$

This map is trace-nonincreasing,

$$\begin{aligned}\mathcal{E}^\dagger\,(\mathbb{1}_{BY}) &= 2^{-\lambda_2}\mathbb{1}_A\otimes\mathcal{T}^\dagger\,(\mathrm{tr}_B\,\mathbb{1}_{BY})\\ &= 2^{-\lambda_2}\mathbb{1}_A\otimes\mathcal{T}^\dagger\left(2^{\lambda_2}\mathbb{1}_Y\right)\leqslant\mathbb{1}_{AX}\ ,\end{aligned}$$

and subunital,

$$\begin{aligned}\mathcal{E}\,(\mathbb{1}_{AX}) &= 2^{-\lambda_2}\mathbb{1}_B\otimes\mathcal{T}\,(\mathrm{tr}_A\,\mathbb{1}_{AX})\\ &= 2^{-\lambda_2}\mathbb{1}_B\otimes\mathcal{T}\left(2^{\lambda_1}\mathbb{1}_X\right)\leqslant\mathbb{1}_{BY}\ ,\end{aligned}$$

since $\lambda=\lambda_1-\lambda_2$ and $\mathcal{T}$ is $2^{-\lambda}$-subunital. Also,

$$\begin{aligned}\mathcal{E}\left(2^{-\lambda_1}\mathbb{1}_A\otimes\sigma_X\right) &= 2^{-\lambda_2}\mathbb{1}_B\otimes\mathcal{T}\left(\mathrm{tr}_A\left(2^{-\lambda_1}\mathbb{1}_A\otimes\sigma_X\right)\right)\\ &= 2^{-\lambda_2}\mathbb{1}_B\otimes\mathcal{T}\,(\sigma_X) = 2^{-\lambda_2}\mathbb{1}_B\otimes\rho_Y\ .\end{aligned}$$

By Prop. 11, we eventually have

$$2^{-\lambda_1}\mathbb{1}_A\otimes\sigma_X\succ_w 2^{-\lambda_2}\mathbb{1}_B\otimes\rho_Y\ .$$

□

Remark 16. *A trace-nonincreasing, $2^{-\lambda}$-subunital completely positive map $\mathcal{T}_{X\to Y}$ can always be written as in Eq. (A4) for a sub-unital trace-nonicreasing completely positive map $\mathcal{E}_{AX\to BY}$, which itself can always be written as projections of a unital map $\mathcal{E}_{CZ\to CZ}$ (see text of the previous proof, and Prop. 12).*

Conversely, for any unital map $\mathcal{E}_{CZ\to CZ}$ with $\mathcal{E}\left(2^{-\lambda_1}\mathbb{1}\otimes\sigma_X\right) = 2^{-\lambda_2}\mathbb{1}\otimes\rho_Y$, in particular for any noisy operation in our framework, the map $\mathcal{T}$ obtained by Eq. (A4) is trace-nonincreasing and $2^{-\lambda}$-subunital.

In particular, for our purposes of optimizing λ over all possible processes of our framework with an additional condition to the channel carrying out the process (namely to preserve correlations between our system X and the reference system R), we may impose that condition directly on the channel $\mathcal{T}$ to obtain an upper bound on λ.

3. Properties for quantum states

We will consider in this section some useful properties of lambda-majorization in the case where we consider normalized states σ, ρ. Here, weak majorization automatically implies (regular) majorization because $\mathrm{tr}\,\sigma = \mathrm{tr}\,\rho = 1$.

In this section, let $\sigma \in \mathscr{S}_=(\mathscr{H}_X)$ and $\rho \in \mathscr{S}_=(\mathscr{H}_Y)$.

Proposition 17 (Lambda-Majorizing a Pure State)**.** *For any pure state $|0\rangle \in \mathscr{H}_Z$, we have $\sigma \xrightarrow{\lambda} |0\rangle\langle 0|$ if and only if $\operatorname{rank}\sigma \leqslant 2^{-\lambda}$ (obviously λ has to be negative or zero). Equivalently, $\sigma \succ \frac{1}{n}\mathbb{1}_n$ if and only if $\operatorname{rank}\sigma \leqslant n$.*

Proof of Prop. 17. Assume first that $\sigma \xrightarrow{\lambda} |0\rangle\langle 0|$. Here $\mathscr{H}_Y$ is the one-dimensional space spanned by $|0\rangle$, and take $\mathscr{H}_X$ the subspace on which σ has its support. By Prop. 9 there exists a single-row matrix ${T_i}^k$ satisfying ${T_i}^k \geqslant 0$, $\sum_i {T_i}^k = T_{i=1}^k \leqslant 1\ \forall k$, $\sum_k {T_i}^k \leqslant 2^{-\lambda}$ such that $1 = \lambda_{i=1}(|0\rangle\langle 0|) = \sum_k T_{i=1}^k \lambda_k(\sigma)$. We also have $\lambda_k(\sigma) \neq 0$ because σ has nonzero eigenvalues in $\mathscr{H}_X$. Then $\sum_k T_{i=1}^k \lambda_k(\sigma) = 1 = \sum_k \lambda_k(\sigma)$ implies $T_{i=1}^k = 1\ \forall k$. That is, the condition $\sum_k T_{i=1}^k \leqslant 2^{-\lambda}$ forces $T_{i=1}^k$ to have at most $2^{-\lambda}$ elements, i.e. the rank of σ may not exceed $2^{-\lambda}$.

The converse holds because any state majorizes a uniform state of the same rank. □

Proposition 18 (Condition on Support Sizes for Lambda-Majorization)**.** *If $\sigma \xrightarrow{\lambda} \rho$, then $\operatorname{rank}\sigma \leqslant 2^{-\lambda}\operatorname{rank}\rho$.*

Proof of Prop. 18. Notice that $\rho \succ \frac{1}{\operatorname{rank}\rho}\mathbb{1}_{\operatorname{rank}\rho}$, and thus $\sigma \xrightarrow{\lambda} \frac{1}{\operatorname{rank}\rho}\mathbb{1}_{\operatorname{rank}\rho}$. Then, by Prop. 8 we have

$$\sigma \xrightarrow{\lambda-\log\operatorname{rank}\rho} |0\rangle\langle 0| \ ;$$

it remains to apply Prop. 17. □

Proposition 19 (Being Lambda-Majorized by a Pure State)**.** *Let the state ρ have maximum eigenvalue $\lambda_{\max}(\rho)$. For any pure state $|0\rangle$, we have $|0\rangle\langle 0| \xrightarrow{\lambda} \rho$ if and only if $\lambda_{\max}(\rho) \leqslant 2^{-\lambda}$. Equivalently, $\frac{1}{n}\mathbb{1}_n \succ \rho$ if and only if $\lambda_{\max}(\rho) \leqslant \frac{1}{n}$.*

Proof of Prop. 19. Let ${T_i}^k$ be as in Prop. 9. Note here k only takes value 1, because we consider $\mathscr{H}_Y$ being the one-dimensional space spanned by $|0\rangle$. Then $\lambda_i(\rho) = \sum_k {T_i}^k \lambda_k(|0\rangle\langle 0|) = {T_i}^{k=1}$ and thus ${T_i}^k = \lambda_i(\rho)$. Then $2^{-\lambda} \geqslant \sum_k {T_i}^k = T_i^{k=1} = \lambda_i(\rho)$ for all i. In particular, $2^{-\lambda} \geqslant \lambda_{\max}(\rho)$.

Conversely, if $\lambda_{\max}(\rho) \leqslant 2^{-\lambda}$, then let ${T_i}^{k=1} = \lambda_i(\rho)$. This matrix T satisfies the conditions in Prop. 9 and thus $|0\rangle\langle 0| \xrightarrow{\lambda} \rho$. □

4. Optimal Lambda Majorization for Normalized States and Relation to Single-Shot Entropy Measures

Define the *absorbed randomness* (or *relative mixedness* [19]) of a transition from σ to ρ as the maximal amount of randomness that you can get rid of, or the minimal amount of randomness that you have to generate, in a noisy operation process:

$$R(\sigma \to \rho) = \sup\left\{\lambda : \sigma \xrightarrow{\lambda} \rho\right\} . \tag{A6}$$

Recent work has shown that this measure is relevant for the amount of extractable work of processes acting on arrays of Szilard boxes [19].

The absorbed randomness has some tight relations to single-shot entropy measures, which we present here. These are reformulations of results shown in [17, 18].

Proposition 20. *The absorbed randomness defined above satisfies the following bounds.*

$$H_{\min}(\rho) - H_0(\sigma) \leqslant R(\sigma \to \rho) \leqslant H_0(\rho) - H_0(\sigma) .$$

Proposition 21. *If $|0\rangle$ denotes any pure state, then the following relations hold:*

$$R(|0\rangle \to \rho) = H_{\min}(\rho) , \tag{A7}$$
$$R(\sigma \to |0\rangle) = -H_0(\sigma) . \tag{A8}$$

Similar explicit values can be obtained in the case where either the initial state or the target state is mixed.

Proposition 22. *If $\frac{\mathbb{1}_n}{n}$ denotes the fully mixed state on $\log n$ qubits, then:*

$$R(\tfrac{\mathbb{1}_n}{n} \to \rho) = H_{\min}(\rho) - \log n , \tag{A9}$$
$$R(\sigma \to \tfrac{\mathbb{1}_n}{n}) = \log n - H_0(\sigma) . \tag{A10}$$

Proof of Prop. 20. Lower bound: Let $\lambda_1 = H_{\min}(\rho) = -\log\lambda_{\max}(\rho)$ and $\lambda_2 = H_0(\sigma) = \log\operatorname{rank}\sigma$. By Proposition 19, we have $2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}} \succ \rho$ and by Proposition 17, $\sigma \succ 2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}}$. The majorization carries over to the tensor product, $2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}} \otimes \sigma \succ 2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}} \otimes \rho$, and $\lambda_1 - \lambda_2$ is a valid maximization candidate for (A6).

Upper bound: Let $\lambda = R(\sigma \to \rho)$ satisfying $\sigma \xrightarrow{\lambda} \rho$. Proposition 18 immediately yields $2^\lambda \leqslant \frac{\operatorname{rank}\rho}{\operatorname{rank}\sigma}$, and

$$R(\sigma \to \rho) = \lambda \leqslant \log\operatorname{rank}\rho - \log\operatorname{rank}\sigma .$$

Recalling the definition of the Rényi-0 entropy $H_0(\sigma) = \log\operatorname{rank}\sigma$ yields the required upper bound. □

Proof of Prop. 21. Equation (A8) follows from the bounds of Proposition 20, which become tight in this special case. Equality (A7) is a direct consequence of Prop. 19. □

Proof of Prop. 22. The bounds of Proposition 20 become tight for (A10). Equality (A9) is again a consequence of Prop. 19, recalling Prop. 8 which allows us to write $|0\rangle\langle 0| \xrightarrow{\lambda+\log n} \rho$ instead of $\frac{\mathbb{1}_n}{n} \xrightarrow{\lambda} \rho$. □

Appendix B: Derivation of the Main Result: Formulation as Semidefinite Program

Let $\mathscr{H}_X$ be a quantum system in the state σ_X. Let $\mathscr{H}_R$ be an additional quantum system and let $|\sigma\rangle_{XR}$ be a purification of σ_X.

Suppose we want to bring the system X into a given state ρ_{XR} with a lambda-majorization (here ρ_{XR} is not necessarily pure; giving the joint state with R allows us to specify which correlations we want to preserve). The task is then the following.

Task. Find the best (maximal) λ, such that there exists a completely positive, $2^{-\lambda}$-subunital, trace-nonincreasing map $\mathcal{T}_{X\to X'}$ satisfying $\mathcal{T}_{X\to X'}(\sigma_{XR}) = \rho_{X'R}$.

In other words, we would like to find the trace non-increasing channel that satisfies $\mathcal{T}_{X\to X'}(\sigma_{XR}) = \rho_{X'R}$, that has the smallest possible $\|\mathcal{T}_{X\to X'}(\mathbb{1}_X)\|_\infty$.

This problem can be formulated as a semidefinite program in terms of the variables α (defined as $\alpha = 2^{-\lambda}$) and $\mathcal{T}_{X\to X'}$ (through its Choi-Jamiolkowski map $T_{XX'}$). (See [60, 61] for a introduction to SDPs in a style similar to what we use here.)

<u>Primal</u>

`minimize:` α

`subject to:`

$$\mathcal{T}_{X\to X'}(\mathbb{1}_X) \leqslant \alpha \mathbb{1}_{X'} \quad : \omega_{X'} \tag{B1a}$$

$$\mathcal{T}^\dagger_{X\leftarrow X'}(\mathbb{1}_{X'}) \leqslant \mathbb{1}_X \quad : X_X \tag{B1b}$$

$$\mathcal{T}_{X\to X'}(\sigma_{XR}) = \rho_{X'R}\ . \quad : Z_{X'R} \tag{B1c}$$

<u>Dual</u>

`maximize:`

$$\operatorname{tr}(Z_{X'R}\,\rho_{X'R}) - \operatorname{tr} X_X$$

`subject to:`

$$\operatorname{tr}\omega_{X'} \leqslant 1 \tag{B2a}$$

$$\operatorname{tr}_R\left[\sigma_{XR}^{t_X} Z_{X'R}\right] \leqslant \mathbb{1}_X \otimes \omega_{X'} + X_X \otimes \mathbb{1}_{X'}\ . \tag{B2b}$$

Note that since the channel does not touch σ_R, we must necessarily have $\sigma_R = \rho_R$. Let E be an environment that purifies the output state as $\rho_{X'RE}$. As two purifications with the same reduced state on R, the two states σ_{XR} and $\rho_{X'R}$ must be related by an isometry $V_{X\to X'E}$ as $\rho_{X'RE} = V_{X\to X'E}\,\sigma_{XR}\,V^\dagger$. We can choose $V_{X\to X'E}$ to be a partial isometry such that $VV^\dagger = \hat{\Pi}_{X'E}$, the projector on the support of $\rho_{X'E}$, and $V^\dagger V = \Pi_X$, the projector on the support of σ_X.

Now, define $\mathcal{T}$ by its Stinespring dilation

$$\mathcal{T}_{X\to X'}(\cdot) = \operatorname{tr}_E\left[V_{X\to X'E}\,(\cdot)\,V^\dagger\right]\ , \tag{B3}$$

and let $\alpha = \|\mathcal{T}(\mathbb{1}_X)\|_\infty$. We will show that this choice of variables is feasible and optimal, and will derive a more explicit value of α.

Condition (B1a) is satisfied by definition and (B1b) because V is a partial isometry. Also, verifying condition (B1c),

$$\mathcal{T}_{X\to X'}(\sigma_{XR}) = \operatorname{tr}_E\left[V_{X\to X'}\,\sigma_{XR}V^\dagger\right] = \operatorname{tr}_E\,\rho_{X'RE} = \rho_{X'R}\ . \tag{B4}$$

Now calculate

$$\alpha = \|\mathcal{T}(\mathbb{1}_X)\|_\infty = \|\operatorname{tr}_E VV^\dagger\|_\infty = \|\operatorname{tr}_E \hat{\Pi}_{X'E}\|_\infty = \max_{\tau_{X'}} \operatorname{tr}\left[\hat{\Pi}_{X'E}\,\tau_{X'}\right] = 2^{H_0(E|X')_\rho}\ . \tag{B5}$$

We will now show that this value is optimal by exhibiting a solution to the dual problem that achieves the same value. Let $\omega_{X'} = \tau_{X'}$ be the optimal $\tau_{X'}$ for the definition of $H_0(E|X')$ as in (B5), let $Z_{X'R} = \sigma_R^{-1} \otimes \omega_{X'}$ and let $X_X = 0$. This choice is feasible since condition (B2a) is automatically satisfied and condition (B2b) becomes

$$\begin{aligned}\operatorname{tr}_R\left[\sigma_{XR}^{t_X} Z_{X'R}\right] &= \operatorname{tr}_R\left[\sigma_{XR}^{t_X}\cdot\rho_R^{-1}\otimes\omega_{X'}\right]\\ &= \operatorname{tr}_R\left[\Phi_{X|R}^{t_X}\otimes\omega_{X'}\right]\\ &= \Pi_X^{t_X}\otimes\omega_{X'} \leqslant \mathbb{1}_X\otimes\omega_{X'}\ ,\end{aligned} \tag{B6}$$

where $\Phi_{X|R}$ is a maximally entangled state on the supports of σ_X and σ_R. Let $\rho_{X'RE}$ and $V_{X\to X'E}$ be defined as before. The value achieved by this choice of dual variables is then

$$\operatorname{tr}\left[Z_{X'R}\,\rho_{X'R}\right] = \operatorname{tr}\left[\sigma_R^{-1}\otimes\omega_{X'}\cdot\rho_{X'R}\right] \tag{B7}$$

$$= \operatorname{tr}\left[\sigma_R^{-1}\otimes\omega_{X'}\cdot V_{X\to X'E}\,\sigma_{XR}V^\dagger\right] \tag{B8}$$

$$= \operatorname{tr}\left[\omega_{X'}\cdot V_{X\to X'E}\,\Phi_{X|R}V^\dagger\right]$$

$$= \operatorname{tr}\left[\omega_{X'}\hat{\Pi}_{X'E}\right] = 2^{H_0(E|X')_\rho}\ . \tag{B9}$$

From this, we conclude that the optimal λ for this problem is

$$\lambda_{\mathrm{opt}} = -H_0(E|X')_\rho\ . \tag{B10}$$

where $\rho_{X'RE}$ is a purification of $\rho_{X'R}$.

We note also that this gives the optimal amount of extracted work. Of course, any $\lambda \leqslant \lambda_{\mathrm{opt}}$ also is a solution.

Appendix C: Rényi-zero entropy of the W state

Let S and M be two qubits in the state $\rho_{SM} = \frac{1}{3}|00\rangle\langle 00|_{SM} + \frac{2}{3}|\Psi^+\rangle\langle\Psi^+|$ (where $|\Psi^+\rangle$ is the Bell state $|\Psi^+\rangle = \frac{1}{\sqrt{2}}\left[|01\rangle + |10\rangle\right]$). Written out explicitely in the basis $\{|0\rangle, |1\rangle\}$,

$$\rho_{SM} = \begin{pmatrix} 1/3 & & & \\ & 1/3 & 1/3 & \\ & 1/3 & 1/3 & \\ & & & 0 \end{pmatrix}\ .$$

(Empty entries are zero.)

The projector on its support is

$$\Pi_{SM} = \begin{pmatrix} 1 & & & \\ & 1/2 & 1/2 & \\ & 1/2 & 1/2 & \\ & & & 0 \end{pmatrix} .$$

We would like to compute the quantity

$$2^{H_0(S|M)_\rho} = \max_{\sigma_M \text{ dens. op.}} \operatorname{tr} \Pi_{SM} \sigma_M .$$

Let $\sigma_M = \begin{pmatrix} s_1 & s_2^* \\ s_2 & 1-s_1 \end{pmatrix}$; then

$$\begin{aligned} \operatorname{tr}\left[\Pi_{SM}(\mathbb{1}_S \otimes \sigma_M)\right] &= s_1 + \frac{1}{2}(1-s_1) + \frac{1}{2}s_1 \\ &= \frac{1}{2} + s_1 . \end{aligned} \quad \text{(C1)}$$

Under the constraint $0 \leqslant s_1 \leqslant 1$, this expression is clearly maximized when $s_1 = 1$, yielding the value

$$H_0(S|M)_\rho = \log \frac{3}{2} .$$

---

[1] R. Landauer, IBM Journal of Research and Development **5**, 183 (1961), ISSN 0018-8646.
[2] C. H. Bennett, International Journal of Theoretical Physics **21**, 905 (1982).
[3] C. H. Bennett, Studies in History and Philosophy of Modern Physics **34**, 501 (2003).
[4] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki, Physical Review Letters **89**, 180402 (2002), ISSN 0031-9007, 0112074.
[5] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral, Nature **474**, 61 (2011), ISSN 1476-4687.
[6] C. H. Bennett, IBM Journal of Research and Development **17**, 525 (1973), ISSN 0018-8646.
[7] L. Szilard, Zeitschrift für Physik **53**, 840 (1929), ISSN 1434-6001.
[8] R. W. Keyes and R. Landauer, IBM Journal of Research and Development **14**, 152 (1970), ISSN 0018-8646.
[9] K. Maruyama, F. Nori, and V. Vedral, Reviews of Modern Physics **81**, 1 (2009), ISSN 0034-6861.
[10] B. Piechocinska, Physical Review A **61**, 062314 (2000), ISSN 1050-2947.
[11] L. Schulman and U. Vazirani, in *Proceedings of the thirty-first annual ACM symposium on Theory of computing* (ACM, 1999), pp. 322–329.
[12] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, Open Systems & Information Dynamics **11**, 205 (2004), ISSN 1230–1612.
[13] M. Horodecki, P. Horodecki, and J. Oppenheim, Physical Review A **67**, 10 (2003), ISSN 1050-2947.
[14] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, ArXiv e-prints (2011), 1111.3882.
[15] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, International Journal of Theoretical Physics **39**, 2717 (2000).
[16] D. Janzing, *Computer science approach to quantum control* (Universitätsverlag Karlsruhe, Karlsruhe, 2006), ISBN 3-86644-083-9.
[17] D. Egloff, Master's thesis, ETH Zürich (2010).
[18] O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral, New Journal of Physics **13**, 53015 (2011), ISSN 1367-2630.
[19] D. Egloff, O. C. O. Dahlsten, R. Renner, and V. Vedral, ArXiv e-prints (2012), 1207.0434.
[20] J. Åberg, ArXiv e-prints (2011), 1110.6121.
[21] M. Horodecki and J. Oppenheim, ArXiv e-prints (2011), 1111.3834.
[22] H. Joe, Journal of Mathematical Analysis and Applications **148**, 287 (1990), ISSN 0022247X.
[23] E. Ruch, R. Schranner, and T. H. Seligman, Journal of Mathematical Analysis and Applications **76**, 222 (1980), ISSN 0022247X.
[24] E. Ruch, R. Schranner, and T. H. Seligman, The Journal of Chemical Physics **69**, 386 (1978), ISSN 00219606.
[25] C. A. Mead, The Journal of Chemical Physics **66**, 459 (1977), ISSN 00219606.
[26] R. Renner, Phd thesis, ETH Zürich (2005), arXiv:quant-ph/0512258.
[27] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Transactions on information theory **56**, 4674 (2010).
[28] M. Tomamichel, Ph.D. thesis, ETH Zurich (2012), arXiv:1203.2142.
[29] R. König, R. Renner, and C. Schaffner, IEEE Transactions on Information Theory **55**, 4337 (2009), ISSN 0018-9448.
[30] N. Datta and R. Renner, IEEE Transactions on Information Theory **55**, 2807 (2009), ISSN 0018-9448.
[31] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics (Springer, 1997).
[32] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, New York, 1985).
[33] A. Uhlmann, Reports on Mathematical Physics **1**, 147 (1970), ISSN 00344877.
[34] A. Uhlmann, Wiss. Z. Karl-Marx-Univ. Leipzig, Math.-Naturwiss. **20**, 633 (1971).
[35] A. Uhlmann, Wiss. Z. Karl-Marx-Univ. Leipzig, Math.-Naturwiss. **21**, 421 (1972).
[36] A. Uhlmann, Wiss. Z. Karl-Marx-Univ. Leipzig, Math.-

Naturwiss. **22**, 139 (1973).
[37] U. Haagerup and M. Musat, Communications in Mathematical Physics **303**, 555 (2011), ISSN 0010-3616.
[38] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1: Contributions to the Theory of Statistics* (1960), pp. 547–561.
[39] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, IEEE Transactions on Information Theory **57**, 5524 (2011), ISSN 0018-9448.
[40] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
[41] M. Tomamichel, R. Colbeck, and R. Renner, IEEE Transactions on information theory **55**, 5840 (2009).
[42] H. E. D. Scovil and E. O. Schulz-DuBois, Phys. Rev. Lett. **2**, 262 (1959).
[43] J. E. Geusic, E. O. Schulz-DuBios, and H. E. D. Scovil, Phys. Rev. **156**, 343 (1967).
[44] R. Alicki, Journal of Statistical Physics **20**, 671 (1979), ISSN 0022-4715.
[45] J. Howard, Nature **389**, 561 (1997).
[46] E. Geva and R. Kosloff, The Journal of chemical physics **97**, 4398 (1992).
[47] P. Hänggi and F. Marchesoni, Rev. Mod. Phys. **81**, 387 (2009).
[48] A. E. Allahverdyan and T. M. Nieuwenhuizen, Phys. Rev. Lett. **85**, 1799 (2000).
[49] T. Feldmann and R. Kosloff, Phys. Rev. E **73**, 025107 (2006).
[50] J. Baugh, O. Moussa, C. Ryan, A. Nayak, and R. Laflamme, Nature **438**, 470 (2005).
[51] J. Gemmer, M. Michel, M. Michel, and G. Mahler, *Quantum thermodynamics: Emergence of thermodynamic behavior within composite quantum systems* (Springer Verlag, 2009).
[52] S. Popescu, A. Short, and A. Winter, Nature Physics **2**, 754 (2006).
[53] N. Linden, S. Popescu, and P. Skrzypczyk, Physical Review Letters **105**, 130401 (2010).
[54] N. Linden, S. Popescu, A. Short, and A. Winter, New Journal of Physics **12**, 055021 (2010).
[55] C. Gogolin, M. Müller, and J. Eisert, Physical Review Letters **106**, 40401 (2011).
[56] S. Trotzky, Y. Chen, A. Flesch, I. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, Nature Physics **8**, 325 (2012).
[57] A. Hutter and S. Wehner, ArXiv e-prints (2011), 1111.3080.
[58] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications* (Springer, 2010), ISBN 0387400877.
[59] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge Mathematical Library (Cambridge University Press, 1952), ISBN 9780521358804.
[60] A. Barvinok, *A Course in Convexity*, vol. 54 of *Graduate Studies in Mathematics* (American Mathematical Society, 2002), ISBN 0-8218-2968-8.
[61] J. Watrous, *Theory of Quantum Information—Lecture notes from Fall 2008* (Online Lecture Notes, 2008), www.cs.uwaterloo.ca/~watrous/quant-info/.